RICE UNIVERSITY

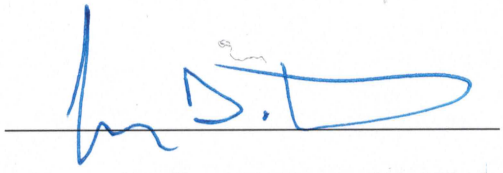
Charge Pairs and Mutations in Collagen Mimetic Peptides

by

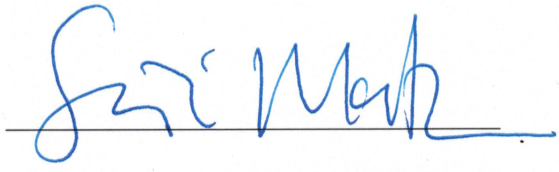
Katherine Ann Clements

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
Doctor of Philosophy

APPROVED, THESIS COMMITTEE



Jeffrey D. Hartgerink, Chair
Professor of Chemistry



Seiichi P.T. Matsuda
Dean of Graduate and Postdoctoral Studies
E. Dell Butcher Professor of Chemistry and
Professor of Biochemistry and Cell Biology

George Phillips
Ralph and Dorothy Looney Professor of
Biochemistry and Cell Biology
Professor of Chemistry

HOUSTON, TX
MAY 2017

Abstract

Charge Pairs and Mutations in Collagen Mimetic Peptides

by

Katherine Ann Clements

This thesis will present insights into natural type I collagen based upon studies with collagen mimetic peptides. Natural collagen is a fibrous protein with challenging properties, such as solubility, which makes it difficult to study through analytical methods such as nuclear magnetic resonance spectroscopy. Collagen mimetic peptides are a flexible and controlled model system that imitates natural type I collagen structure, a triple helix, on a smaller scale, from over 1000 amino acids to around 30 amino acids. Chapter one will explore the development of collagen mimetic peptides and review the current design principle to form composition and register specific triple helices. Chapter two will illustrate the power of charge pairing in the triple helix with the creation of a triple helix with an extended offset through axial charge pairing. This is the first triple helix with an extended offset, proving that the offset triple helix is possible, even when it is not incorporated into a fiber. Chapters three and four will model mutations from Osteogenesis Imperfecta, a type I collagen disease, in collagen mimetic peptides. Osteogenesis Imperfecta results from mutations of the requisite glycines in the Xaa-Yaa-Gly repeat of collagen. Mutations in both the A and B chains of an AAB triple helix will be created, and their impact on the composition, stability, and structure of the triple helix will be examined through circular dichroism and nuclear magnetic resonance experiments. The results of those experiments will be used to create molecular models of the mutated triple helices. This series of studies gives insight into the forces in the triple

helix: the destabilizing force of a mutation, the stabilizing force of hydrogen bonds, and the stabilizing, yet constricting, force of axial charge pairs. In these model systems, the mutation was able to direct the composition and register of the triple helix, where the triple helix with the lowest number of mutations would fold, even if it is not the most stabilized via charge pairing.

Acknowledgements

So many people have supported or assisted me in that marathon that is this thesis, more than I can thank in this one page. First I must thank my advisor, Prof. Hartgerink. He has supported me through the past five years, through all the successful and unsuccessful experiments, and kept pushing me to be the best scientist I could. I couldn't ask for a better advisor for me. I would also like to thank the other professors in the Chemistry department for all their constructive comments and criticism over the past five years. I would especially like to thank Profs. Matsuda and Phillips for their helpful advice throughout my doctorate, as well as for serving on my committee.

I am grateful to all the members of the Hartgerink Lab, each of them has been instrumental in my success. I am eternally grateful to Abhishek for all his patience when training me, and answering all of my questions. Marci was always the voice of reason, helping me put problems in perspective, which I appreciate so much. And Amanda, I will miss all of our idle conversations, conference room debriefings, and celebrations of each other's success. All of you have made the highs of my doctorate sweeter, and lessened the hurt of the lows.

I must thank my friends here in Houston, especially Scott, Sam, and Andrew for nights when I forget about research. I thank my family, for being a constant source of support and confidence. Lastly, my husband Tom for being my constant companion and confidant, I would never be where I am now without him pushing me to be my best. These last five years in Houston with you have been challenging but amazing, and I can't wait for our next adventure.

Table of Contents

List of Figures ..... **vi**

List of Tables ..... **xi**

List of Schemes ..... **xii**

Abbreviations ..... **xiii**

Chapter 1: Introduction **1**
1.1 Collagen Structure and Sequence 1
1.2 Collagen Mimetic Peptides 3
1.3 Design of Collagen Mimetic Triple Helices 4
1.3.1 Single Amino Acid Substitutions 6
1.3.2 Paired Amino Acid Substitutions 7
1.3.3 Charge Pairing 9
1.4 Creating Composition and Register Specific Triple Helices 10
1.4.1 ABC Triple Helices 10
1.4.2 AAB Triple Helices 11
1.4.3 Hydroxyproline and Proline Effects 12
1.4.4 Design Principles 14
1.5 Applications of Collagen Mimetic Peptides 14
1.5.1 Osteogenesis Imperfecta 15
1.5.2 Fiber Forming Peptides 16
1.5.3 Alternate Supramolecular Structures 17
1.5.4 Biological Applications 18
1.6 Conclusions and Preview of Thesis 19
1.7 References 20

Chapter 2: Rational Design of a Non-canonical "Sticky-Ended" Collagen Triple Helix **25**
2.1 Introduction 25
2.2 NMR and CD Analysis of ABC-1 and ABC-2 27
2.2.1 Direct Evidence of Offset 33
2.3 Conclusions 35
2.4 Experimental 37
2.4.1 Peptide Synthesis and Purification 37
2.4.2 Preparation of Samples for CD and NMR 39
2.4.3 Circular Dichroism Analysis 42
2.4.4 Acquisition and Processing Parameters in NMR Spectroscopy 43
2.5 References 45

Chapter 3: AAB Heterotrimers to Investigate Mutations of the B Chain **47**
3.1 Introduction 47
3.2 Characterization of Mutant Helices 49

3.2.1 Base System....49
3.2.2 Mutant Systems....52
3.2.2.1 Alanine....52
3.2.2.2 Serine....54
3.2.2.3 Valine....55
3.2.2.4 Aspartate....56
3.2.2.5 Arginine....57
3.3 Structural Disruptions around Mutation Site....57
3.4 Comparisons to other OI Models....61
3.5 Conclusions....63
3.6 Experimental....66
3.6.1 Peptide Synthesis....66
3.6.2 Peptide Purification....67
3.6.3 Sample Preparation....67
3.6.4 NMR Spectroscopy....67
3.6.5 CD Spectroscopy....68
3.6.6 Molecular Modeling....69
3.7 References....69

Chapter 4: Comparison of Osteogenesis Imperfecta Mutations in the A, B, and all three chains....72
4.1 Introduction....72
4.1.1 Peptide Design....74
4.2 Characterization of Mutant Triple Helices....75
4.2.1 Base AAB Triple Helix....75
4.2.2 A Chain Mutants....76
4.2.3 Triple Mutant Helices....84
4.3 Conclusions....86
4.4 Experimental....87
4.4.1 Peptide Synthesis....87
4.4.2 Peptide Purification....88
4.4.3 Sample Preparation....89
4.4.4 Circular Dichroism Spectroscopy....89
4.4.5 NMR Spectroscopy....90
4.4.6 Molecular Modeling....91
4.4.7 Helical Twist....91
4.5 References....96

Chapter 5: Conclusions....98

Appendix 1: Modeling Constraints....102

Appendix 2: Supporting NMR Spectra....109

Appendix 3: List of Publications....129

Table of Figures

Chapter 1: Introduction 1

1.1 Sequence (a) and model (b) of a collagen triple helix. The chains are offset by one amino acid and each follow the Xaa-Yaa-Gly repeat. The boxes show those specific cross-sections in the triple helix. (c) Model with space-filling glycine residues showing the proximity of the chains. 2

1.2 A) Sequence with hydrogen bonds shown. The A chain is repeated below the C chain to show those hydrogen bonds, because the triple helix is cylindrical a 2D projection is not adequate to illustrate the topology. B) Model of hydrogen bonds shown as yellow dashed lines in a triple helix. 3

1.3 Possible registers of specific compositions.....4

1.4 All possible compositions and registers formed from three distinct peptide strands. 5

1.5 Sequence of Model systems. Melting points of selected substitutions from Persikov, *et al* 2000.....7

1.6 Sequence of the model system. Melting points of selected paired systems from Chan, *et al* 1997. 8

1.7 Sequence and model illustrating axial(a,c) and lateral(b,d) charge pairs. 9

1.8 Found registers of the $(PKGPOG)_5$ $(EOG)_{10}$ system. Amino acids in axial charge pairs shown in grey, while unpaired amino acids are shown in yellow.....12

1.9 (a)Sequence and charge pairs of hydroxyproline-free triple helix. (b)Thermal denaturation monitored by CD of all possible component triple helices. Figure adapted from Jalan et al, 2013. 13

Chapter 2: Rational Design of a Non-canonical "Sticky-Ended" Collagen Triple Helix 25

2.1 (Top) First-derivative plots of thermal unfolding curves of (a) ABC-1 and (b) ABC-2 overlaid with those of component peptides (shown in black) and (c) overlay of ABC-1 and ABC-2 spectra. (Bottom) $^{1}H,^{15}N$-HSQC NMR spectra of (d) ABC-1 and (e) ABC-2 overlaid with those of component peptides (shown in black) and (f) overlay of ABC-1 and ABC-2 HSQC spectra, demonstrating the similarity of two of the three N−H peaks. Monomer peaks are indicated within the box. 29

2.2 (a) Sequence, (b) ^{15}N-edited-NOESY spectrum, and (c) model of ABC-1 indicating absence of inter-peptide $^{15}NH-^{15}NH$ (red lines) and $^{15}NH-H\alpha$ (blue lines) NOE cross-

peaks due to >6 Å separation between the protons. (d) Sequence, (e) ^{15}N-edited-NOESY spectrum, and (f) model of ABC-2 indicating absence of inter-peptide ^{15}NH−^{15}NH (red lines) and ^{15}NH−Hα (blue lines) NOE cross-peaks due to >6 Å separation between the protons. 30

2.3 Comparison of the hydrogen bonds formed by the ^{15}N-Gly residues (shown in green) in ABC-1 in a non-canonical (a) and ABC-2 in a canonical (b) offset. Peptides in ABC-1 arranged in a hypothetical canonical offset are shown in (c) for comparison. Cyan and red boxes indicate identical chemical environment of ^{15}N-Gly in chains A1 and B of ABC-1 and chains A2 and B of ABC-2. ^{15}N-Gly residues in dissimilar chemical environment are shown within black boxes. 31

2.4 CD thermal melting profile of individual, binary and ternary peptide mixtures used for testing A1, B and C1 32

2.5 ^{1}H,^{15}N-HSQC spectra of ABC-1 one week (left) and three month (right) after mixing A1, B and C1 peptides. The competing heterotrimeric peptide mixture A1C1 is shown in red. The monomer cross peaks are shown in the box.....33

2.6 . a) Position of ^{15}N-Gly labels in ABC-1' shown in green, b) HSQC spectra of ABC-1' acquired on a higher field 750 MHz Bruker cold probe to improve chemical shift dispersion, c) two dimensional ^{1}H-^{1}H version of the 3D-edited-NOESY indicating significant overlap in the amide proton chemical shifts of the three ^{15}N-Gly, d) ^{1}H-^{1}H plane of a three dimensional edited-NOESY experiment at the ^{15}N-chemical shift corresponding to the ^{15}N-Gly of the middle chain. NOEs observed are indicated in model (e).. 34

2.7 HPLC chromatogram (left) and ESI-MS spectra of peptides A1) $WG(POGPKG)_5$ monoisotopic mass$[M+2H]^{2+}$, expected: 1504.8 m/z, observed: 1504.8 m/z., A2) $WG(PKGPOG)_5$ monoisotopic mass$[M+3H]^{3+}$, expected: 1017.5 m/z, observed: 1017.6 m/z and B) $WG(PKGEOG)_5$ monoisotopic mass $[M+2H]^{2+}$, expected: 1584.8 m/z, observed: 1584.8 m/z. 40

2.8 HPLC chromatogram (left panel) and ESI-MS (right panel) of peptides. C1) $WG(EOGPOG)_5$ monoisotopic mass $[M-2H^{-}]^{2-}$, expected: 1545.1 m/z, observed: 1545.3 m/z, C1') $WG(EPGPPG)_5$ monoisotopic mass $[M+2K]^{2+}$, expected: 1505.1 m/z, observed: 1505.4 m/z and C2) $WG(POGEOG)_5$ monoisotopic mass $[M-2H]^{2-}$, expected: 1545.1 m/z, observed: 1545.3 m/z.....41

2.9 Thermal melting profile (A) and their first derivative curves (B) of individual and binary mixtures of all peptides used in this study.....43

2.10 CD wavelength scan (left panel) and thermal melting profile (right panel) of ABC-1 (green) and ABC-2 (red). The wavelength scans of both ABC heterotrimers show similar CD spectra with a positive ~ 225 nm maxima and a negative ~200 nm minimum characteristic of the left-handed poly-proline type II peptide conformation in

the triple helix. A cooperative thermal unfolding is observed for both heterotrimers in the right panel. 44

Chapter 3: AAB Heterotrimers to Investigate Mutations of the B Chain.....47

3.1 Sequence of A and B peptides used. Green represents the mutation site, blue amino acids are positively charged and red amino acids are negatively charged. Boxed amino acids are ^{15}N-isotopically labeled glycines and are numbered for easy correlation to NMR data.....48

3.2 ^{1}H,^{15}N-HSQC spectra of base and mutant systems. Blue traces show the B' mutant peptide alone, red traces show the A peptide alone. Black shows the 2:1 AAB' mixture of the peptides. Peaks labeled M represent monomer, those labeled A represent A homotrimer while peaks labeled G represent the AAB' mutant labels as labeled in Fig. 3.1.....50

3.3 NOESY HSQC of the AAB base peptide.....51

3.4 First derivative of Circular Dichroism melting experiments demonstrating the relative thermal stabilities of the compared systems. For easy comparison, the component peptides AAA and BBB and their combined AAB triple helix are shown in all graphs in red, blue and green respectively. Modified peptides with substitutions away from Glycine are shown in black.....54

3.5 .Figure 3.5. NOESY-HSQC of B-Ala (a) and B-Ser (b). 58

3.6 a.) Ramachandran plot of B-Ala helix compared to Base b.) Chart of interactions lost and gained in B-Ala model. R_1 represents lysine side chain, R_2 represents glutamate side chain, and R_3 represents aspartate side chain. Blue interactions are expected backbone hydrogen bonds that formed, green are expected axial charge pairs that are formed, red are broken hydrogen bonds or charge pairs, and purple are new compensating interactions c) Generated model of B-Ala. Alanine is shown in yellow 59

3.7 a.) Ramachandran plot of B-Ser helix compared to base b.) Chart of interactions lost and gained in B-Ser model. R_1 represents lysine side chain, R_2 represents glutamate side chain, and R_3 represents aspartate side chain. Blue interactions are expected backbone hydrogen bonds that formed, green are expected axial charge pairs that formed, red are broken hydrogen bonds or charge pairs, and purple are compensating interactions that formed. * denotes an interaction far from labeling site that is most likely a product of termini affects, not the mutation. c) Model of B-Ser. Serine shown in yellow.....60

3.8 Triple Helical Radius of the mutated, base, and 1CAG helices 62

3.9 a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3210.55 Observed =3211.48) and b) HPLC trace of the Base A peptide. c) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3062.23 Observed =3061.60) and d) HPLC trace of the Base B peptide. e) ESI-TOF MS of the purified fractions (Predicted

MW $[M+H]^+$ =3079.25 Observed =3078.08) and f) HPLC trace of the B-Ala B peptide. 64

3.10 a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3095.23 Observed =3094.48) and b) HPLC trace of the B-Ser B peptide. c) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3107.12 Observed =3108.11) and d) HPLC trace of the crude B-Val B peptide. e) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3123.08 Observed =3122.80) and f) HPLC trace of the crude B-Asp B peptide 65

3.11 . a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3164.10 Observed =3164.30) and b) HPLC trace of the crude B-Arg B peptide. 66

Chapter 4: Comparison of Osteogenesis Imperfecta Mutations in the A, B, and all three chains.....72

4.1 Sequences, compositions and registers of peptides used. Red amino acids are negatively charged, blue are positively charged. The yellow boxed glycines are ^{15}N isotopically labeled. A is the B-Ala mutant, B is A-Ala mutant which is observed to form, C is A-Ala mutant which was intended but is not observed to fold, D is the 3-Ala mutant. The serine mutants are identical other than the substitution of alanine for serine. 75

4.2 $^{1}H,^{15}N$-HSQC of the peptide systems. Black spectra are the AAB mixture. Blue is only the B strand used in the system. Red is only the A strand used in the system. G1-G5 denote the heterotrimer present and are labeled in the same order as shown in Figure 1. A1-A3 denote the AAA homotrimers. M1-M3 denote the peptide monomers. B-Ala and B-Ser spectra are reproduced from reference 13 for comparison. 77

4.3 Circular Dichroism. First derivatives of the melting plots are shown for clarity. The melting temperature is defined as the minimum of the curve. The Base AAA peptide is shown in all three panels because it is a competing homotrimer in the solution. 78

4.4 NOESY-HSQC of A-Ala (a) and A-Ser (b). 79

4.5 Interactions in (a,b) B-Ala,[13] (c,d) A-Ala, (e,f) 3-Ala. a, c, and e are models of the mutation site. 82

4.6 Interactions in (a,b) B-Ser,[13] (c,d) A-Ser, (e,f) 3-Ser. a,c, and e are models of the mutation site. 83

4.7 Helical twist of the peptide systems. The first (leading) chain of the triple helix is in red, the second (middle) is in green, and the third (lagging) is in blue. The vertical grey area is the mutated region, and the horizontal grey area is the range of helical twist in a POG model. The x-axis is the cross-section of the triple helix, starting where the lagging strand joins the triple helix. 84

4.8 a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^{+}$= 3324.49, Observed 3224.67) and b) HPLC trace of the A-Ala A peptide. c) ESI-TOF MS of the purified fractions (Predicted MW $[M+K+H]^{3+}$= 3279.50, Observed 3279.23) and d) HPLC trace of the A-Ser A peptide. 88

4.9 Panel A illustrate two adjacent triads composed of the three α carbons in each cross section. The displacement between the centroids is shown as d and the radius as r. Panel B shows the two triad centroids overlaid to illustrate the tilt angle between them. In panel C, the triads are viewed from above and have been tilted such that they lie in the same plane. Illustrated are the angles of interest in each respective color green, turquoise, and white. 92

4.10 Illustration of the intersection of the planes created by each triad. 95

Table of Tables

Chapter 2: Rational Design of a Non-canonical "Sticky-Ended" Collagen Triple Helix 25

2.1 Sequences, Abbreviations, and Melting Temperatures of Triple Helices Studied. Bold glycines indicate ^{15}N isotopically labeled amino acids. M indicates a broad melting transistion. 28

2.2 Comparison of the observed ^{15}NH-^{15}NH and ^{15}NH-Hα NOE correlations in ABC-2 to those expected from the six possible registers. Check marks indicate that the expected cross peak is observed while the X marks indicate otherwise. 35

Chapter 3: AAB Heterotrimers to Investigate Mutations of the B Chain.....47

3.1 Alpha homotrimer relaxation values.....52

3.2 Peak volumes from ^{1}H, ^{15}N-HSQC 52

3.3 Summary of experimental properties of the sequences. Melting points are from Circular Dichroism. Relaxation values are from ^{15}N relaxation experiments on the heterotrimeric solutions in numerical order of the label. Peak volumes are from the ^{1}H,^{15}N-HSQC spectra of the mixed heterotrimeric solutions. A more detailed description of the assignment process, additional relaxation values, and TOCSY are given in Appendix 2.....53

3.4 Comparison of lost and gained interactions. Lost hydrogen bonds are backbone hydrogen bonds that would be expected to take place in the Base structure, but no longer are formed in the same way. Compensating interactions are any interactions that are not present in the Base structure.....56

Chapter 4: Comparison of Osteogenesis Imperfecta Mutations in the A, B, and all three chains.....72

4.1 NMR Experimental Results of the different Osteogenesis Imperfecta model systems. Heterotrimer T_m is the melting point of the folded heterotrimer by CD. The R2/R1 value, or relaxation value, is for each isotopic label (G1-G5) in the order from leading to lagging chain. Relative volumes are the relative volume of the peaks when the total peak volume is set to 1 (see experimental procedures for more details). 76

4.2 Results of the modeling experiments. Lost hydrogen bonds are backbone hydrogen bonds that are expected to be found in the model, but are not present. Lost charge pairs are designed axial charge pairs that are not formed in the model. Compensating interactions are hydrogen bonds that are not present in the base structure, but are formed in the model. Formed charge pairs are designed axial charge pairs that are present in the model.....81

Table of Schemes

Chapter 2: Rational Design of a Non-canonical "Sticky-Ended" Collagen Triple Helix 25

2.1 Sequences of non-canonical offset ABC-1 and canonical offset ABC-2 heterotrimers, indicating positions of Lys−Glu axial salt-bridges. Grey highlighted area indicates the core region, which is identical in both triple helices.. 28

Chapter 1: Introduction

Natural type I collagen is a native hydrogel with numerous beneficial medical applications, such as wound healing,[1] a scaffold for growth and regeneration of tissues,[2] and drug delivery.[3] However, natural collagen grafts can have issues with immune response.[4] In this review I will describe the recent work in the field of collagen mimetic peptides with a brief review of natural collagen as it relates to the sequence and properties of mimetic peptides. The many applications of natural collagen will not be discussed in detail, but more information can be found in the reviews cited above.

1.1 Collagen Structure and Sequence

Collagen is a complex protein that lends a range of properties to the extracellular matrix, such as strength and adhesion. This supramolecular complex forms large fibers, which create the scaffold for many tissues in the body. Each collagen fiber is composed of smaller fibrils, which in turn are composed of many individual triple helices. Each triple helix has three protein chains, which wind together to create a superhelical structure.[5] Each of these protein chains have a specific triplet repeat amino acid sequence, Xaa-Yaa-Gly, where Xaa and Yaa can be any amino acid, but every third amino acid must be a glycine or the structure will be disrupted.[6] Glycine is essential because it is packed into the center of the triple helix, with the alpha hydrogens almost touching the adjacent chains. Each glycine can be packed into the center of the triple helix because of a one amino acid offset between the chains of the triple helix (Figure 1.1). This offset allows for one glycine to be present in each cross-section of the triple helix, and therefore packed into the center of that cross-section. Any other amino acid causes disruption due

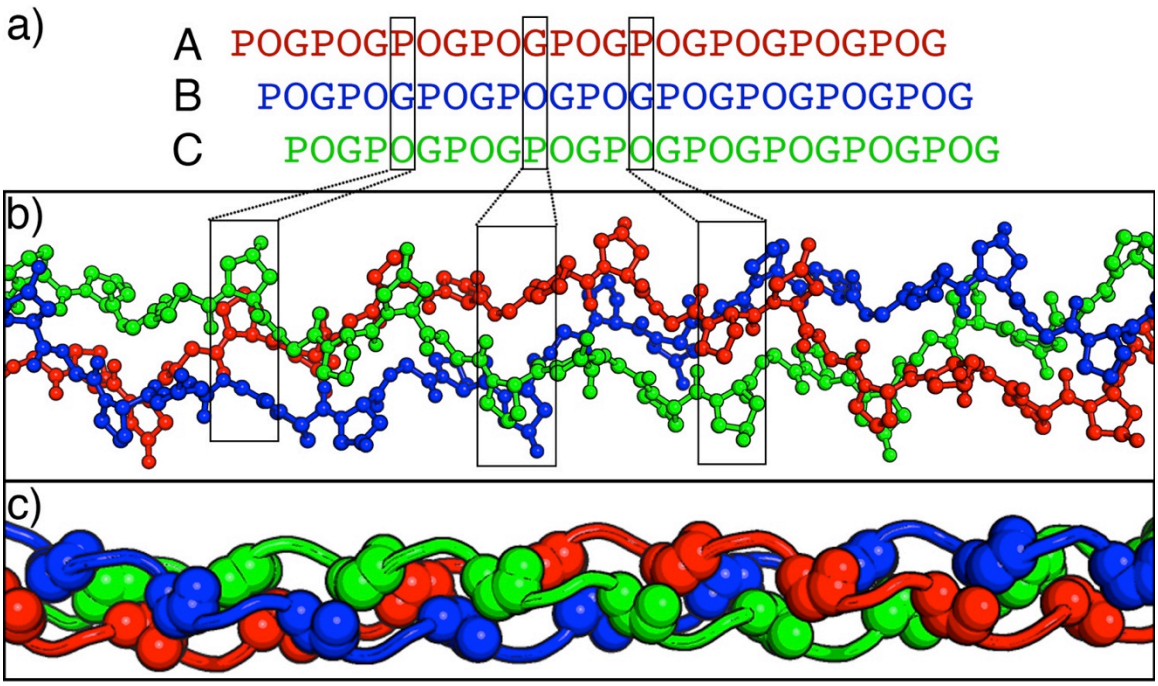


Figure 1.1. Sequence (a) and model (b) of a collagen triple helix. The chains are offset by one amino acid and each follow the Xaa-Yaa-Gly repeat. The boxes show those specific cross-sections in the triple helix. (c) Model with space-filling glycine residues showing the proximity of the chains.

to steric hindrance because it would have a side chain at that position that would disrupt the tight fold of the triple helix.

The Xaa and Yaa positions can be any amino acid, but in natural collagen the Xaa position is dominated by proline (Pro, P), and the Yaa position hydroxyproline (Hyp, O). Hydroxyproline is a post-translational modification of proline in which a hydroxyl group is added to the gamma carbon of the proline ring. The Xaa position is most commonly proline where 32.9% of fiber-forming collagens triplets have proline in the Xaa position, but it is followed by glutamate at 13.0% and alanine at 11.0%.[7] Overall, negatively charged amino acids are more common than positively charged in the Xaa position. The Yaa position is hydroxyproline in 34% of triplets, followed by arginine at 11.4% and

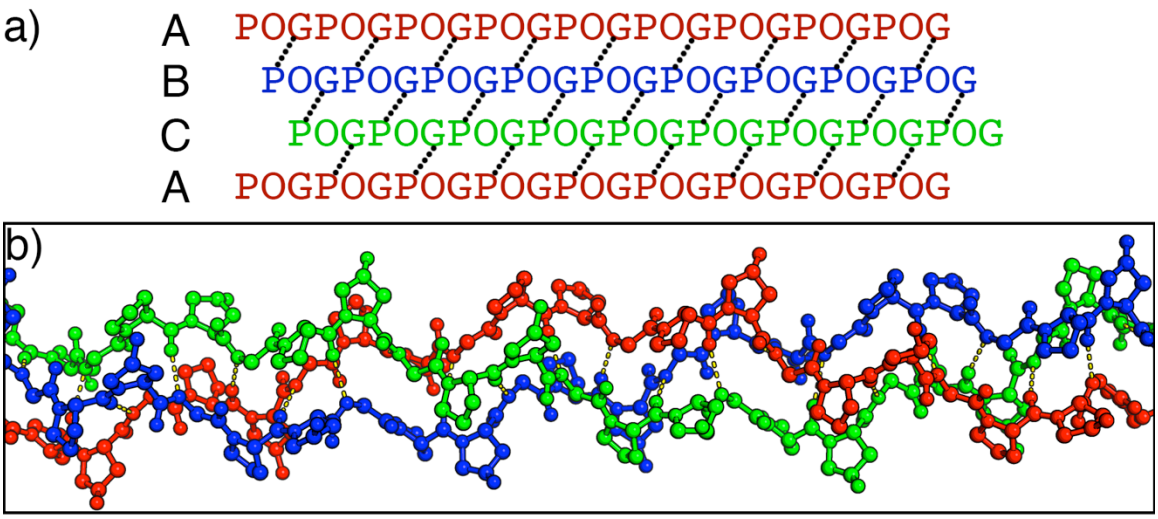


Figure 1.2. A) Sequence with hydrogen bonds shown. The A chain is repeated below the C chain to show those hydrogen bonds, because the triple helix is cylindrical a 2D projection is not adequate to illustrate the topology. B) Model of hydrogen bonds shown as yellow dashed lines in a triple helix.

methionine and lysine both at 9.0%.[7] In the Yaa position positively charged amino acids are more common than negatively charged amino acids.[8,9]

The triple helix is non-covalently linked through hydrogen bonds.[6,10,11] These hydrogen bonds are between a glycine of one chain, and the amino acid in the X position of an adjacent chain (Fig. 1.2). These hydrogen bonds are the major stabilizing force of the triple helix.

1.2 Collagen Mimetic Peptides

Collagen has an active role in various biological processes, acting as a scaffold, an active site for receptors, and a signaling protein.[12] Studying collagen in the lab is vital to understanding how these processes work. However, natural collagen is a fibrous protein with solubility concerns except under harsh conditions. These harsh conditions, such as acid or heat, denature the protein, destroying the structure we wish to study.

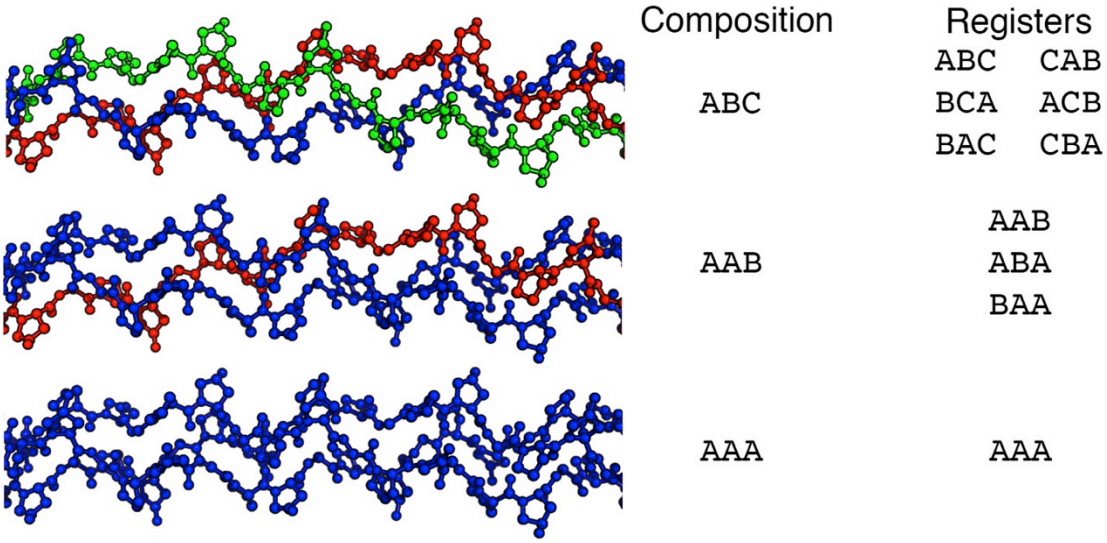


Figure 1.3. Possible registers of specific compositions.

Collagen mimetic peptides replicate the collagen structure on a much smaller scale, ~30 amino acids to the ~1000 amino acids in type I collagen. These shorter peptides are soluble, and can be synthesized with complete sequence control using solid phase peptide synthesis.[13] SPPS allows sequences that are challenging to synthesize in bacterial systems such as modified amino acids,[14] colorimetric labels,[15] or isotopic labels[16] to be incorporated into the synthetic peptides. The peptides self assemble into a triple helix replicating collagen structure.[13] They can be used to study the properties and structure of collagen without the denaturation or complications of the natural system.[16]

1.3 Design of Collagen Mimetic Triple Helices

Collagen mimetic peptides can be designed to form distinct and specific folded triple helices.[17-19] Each triple helix consists of three peptide strands which can have the same sequence, called a homotrimer, or it can have distinct sequences, called a heterotrimer.[20] When a helix has three of the same chains it is called an AAA composition, and there can be AAB or ABC composition as well. Each strand must have the triplet repeat Xaa-Yaa-Gly, but variation in the Xaa and Yaa positions can be used to

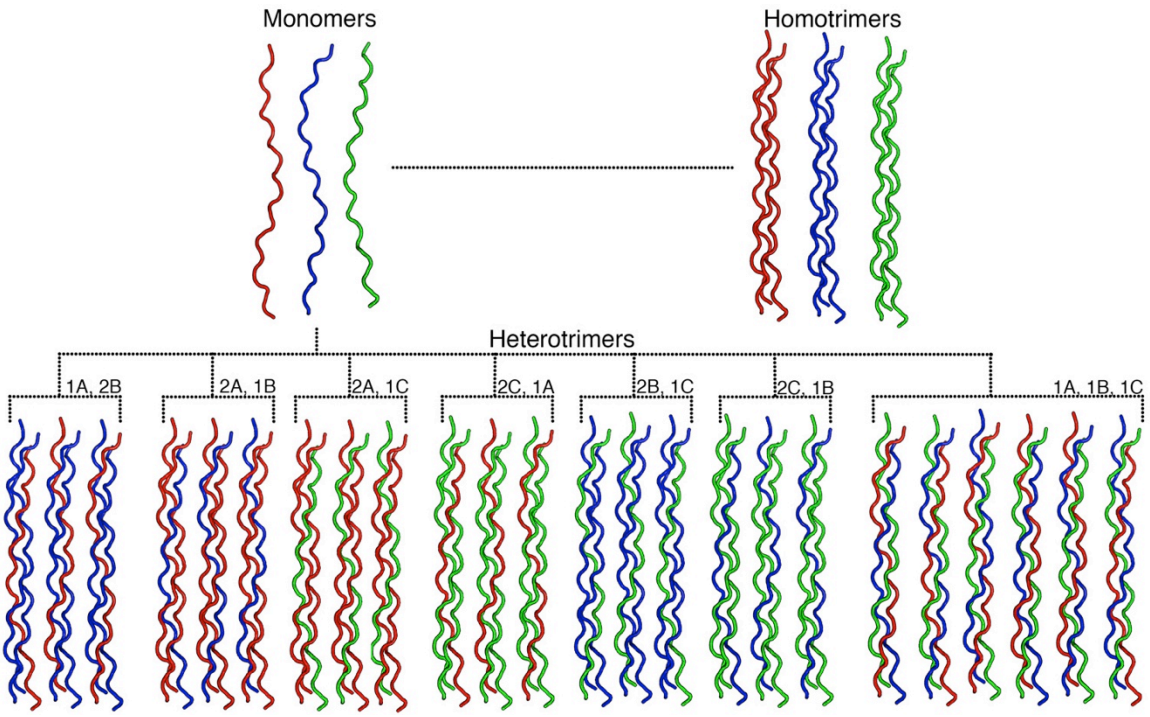


Figure 1.4. All possible compositions and registers formed from three distinct peptide strands.

favor specific triple helices. For example, a particular system has one positively charged peptide strand $(PKG)_{10}$, one negatively charged strand $(DOG)_{10}$, and one neutral strand $(POG)_{10}$. When mixed together these peptides reliably fold into a triple helix in a 1:1:1 ratio.[21]

The one amino acid offset of the triple helix allows for distinction between the chains, and creates differences when the chains are in different orders in the triple helix.[10] For example, in a triple helix with three different peptide strands A, B, and C, different helices are created if the chains fold in the order ABC or BAC. ABC and BAC are called registers, or orders, of the peptides in the ABC composition triple helix, illustrated in Fig. 1.3.[20]

Since these collagen mimetic peptides self assemble there is a complex equilibrium when one, two, or three peptides are mixed together.[20] If one register is very

favorable energetically, it will be the major assembly, but if multiple triple helices are favorable there will be many components to the solution. When three peptides are mixed together there are twenty-seven possible triple helices (Fig. 1.4). A specific composition and register could be required for a disease model or study of binding sites. Designing a triple helix requires consideration of all the alternate triple helices and comparing them to the target.

In a triple helix, the most thermally stable triplet sequence is proline-hydroxyproline-glycine.[7] Any substitution away from these amino acids destabilizes the triple helix.[7] However, simple POG triple helices cannot have a controlled composition and register, nor do they adequately reflect natural collagen, which has a wide range of amino acids.

1.3.1 Single Amino Acid Substitutions

Most amino acids have been studied for their individual impact on the stability of the triple helix. In this study a $(GPO)_3GPY(GPO)_4GG$ host-guest peptide was used where individual substitutions were made in the Yaa position, the peptide was incubated, and the melting point of the resulting triple helix was measured.[7] The most stable amino acids in the Y position were Hyp, Arg, and Met while the least stable were His, Asp, Leu, Asn, Tyr, Phe, and Trp (Fig. 1.5).[7] Computational modeling indicates that stabilization from Arg and Met may be due to noncovalent interaction of the side chains with the backbone of the peptide chain. However, in the least stabilizing amino acids, branching at the delta carbon blocks solvent molecules from accessing the backbone, reducing the stabilization of solvation.[7]

a) GPOGPOGPOGPYGPOGPOGPOGPOGG
GPOGPOGPOGPYGPOGPOGPOGPOGG
GPOGPOGPOGPYGPOGPOGPOGPOGG

b) GPOGPOGPOGXOGPOGPOGPOGPOGG
GPOGPOGPOGXOGPOGPOGPOGPOGG
GPOGPOGPOGXOGPOGPOGPOGPOGG

Substitution	GXO	GPY
Hyp		47.3
Arg	40.6	47.2
Met	38.6	42.6
Ile	38.4	41.5
Gln	40.4	41.3
Glu	42.9	39.7

Figure 1.5. Sequence of Model systems. Melting points of selected substitutions from Persikov, *et al* 2000.

A complementary system $(GPO)_3GPX(GPO)_4$ was used to do the same analysis with amino acids in the Xaa position. In the Xaa position Pro, Glu, Ala, Lys, Arg, Gln, Asp are the most stable (Fig. 1.5). The ionizable side chains of Glu, Lys, Arg, Gln, and Asp are available to hydrogen bonds with the backbone carbonyls or solvent molecules. Residues with branching at the delta carbon are less stable in the Xaa position when compared to the Yaa position. However, the most stabilizing amino acids are the imino acids, proline and hydroxyproline, where proline is the most stabilizing in the Yaa position and hydroxyproline is the most stabilizing in the Xaa position.[7]

1.3.2 Paired Amino Acid Substitutions

The single amino acid substitutions are isolated with only proline, hydroxyproline, and glycine available for interactions. However, with double substitutions additive effects between the amino acids can be explored. The base sequence $(GPO)_3GXY(GPO)_4GG$ was used to test different combination of amino acids selected based upon the frequency of the triplet occurring in natural collagen. The homotrimeric triple helices were folded and the melting point of the helix was found. The triplets GPO, GER, GAR, GDR, GKD, GKE, GEK, GRD, and GRE had the highest melting points in excess of 33 °C (Fig 1.6).[8] Seven of these nine sequences incorporate both a positive and negative amino acid. These charge pairs suggest a stabilizing

a) GPOGPOGPOGXYGPOGPOGPOGPOGG
GPOGPOGPOGXYGPOGPOGPOGPOGG
GPOGPOGPOGXYGPOGPOGPOGPOGG

b) GPOGPOGPOGXYGX'Y'GPOGPOGPOGG
GPOGPOGPOGXYGX'Y'GPOGPOGPOGG
GPOGPOGPOGXYGX'Y'GPOGPOGPOGG

Substitution	T_m °C	Substitution	T_m °C
GPKGEO	47.8	GDKGPO	30.9
GEOGPK	38.0	GDRGPO	37.1
GPKGPE	36.5	GEKGPO	35.0
GPKGDO	47.1	GERGPO	40.4
GPRGDO	39.6	GKEGPO	35.3
GPRGEO	42.8	GKDGPO	35.8

Figure 1.6. Sequence of the model system. Melting points of selected paired systems from Chan, *et al* 1997.

interaction when both a positive and negative amino acid are present in the triple helix when compared to sequences with two uncharged amino acids.

Further studies have refined the positional preferences of the positively and negatively charge amino acids showing a preference of negative residue for the Xaa position and positive amino acids for the Yaa position.[8,22] Furthermore, the distance between the amino acids is also important. Utilizing the model system $(GPO)_3$G-X-Y-G-X'-Y'-$(GPO)_3$GG, where X and X' can be independently substituted, interactions where the amino acids were separated by three amino acids were studied. The guest sequences of GPKGEO and GPKGDO showed surprisingly melting points of 47.8 °C and 47.1 °C, respectively, drastically above the predicted values of 32.4 °C and 29.6 °C.[23] The predicted melting points are calculated from the decrease in melting point from $(GPO)_8$GG when a single lysine and a single glutamic acid is substituted (for GPKGEO). These predicted values should predict the melting point of the helix if there is no interaction or secondary effects of having two amino acids substituted. However, the triple helices were much more stable than the prediction, indicating interaction between these amino acids because they exceed the simple additive prediction of the stability of the triple helix.

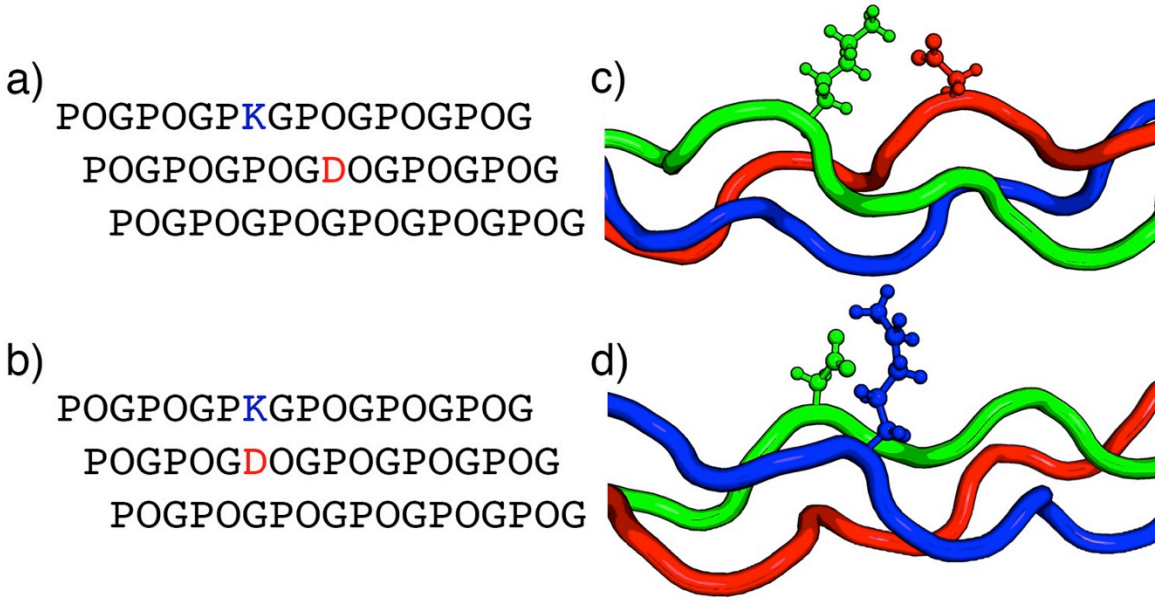


Figure 1.7. Sequence and model illustrating axial(a,c) and lateral(b,d) charge pairs.

1.3.3 Charge Pairing

Interactions of two oppositely charged amino acids in the triple helix are called a charge pairs. The orientation and identity of the amino acids determines the energetic benefit of the pair. Two common orientations are lateral and axial charge pairs.[24] The lateral charge pair consists of a positively charged amino acid that is one position over in the sequence on an adjacent chain from a negatively charged amino acid. The amino acids are aligned in the triple helix, essentially standing out from the helix next to one another, and the positively charged amino acid leans over to interact with the negatively charged side chain (Fig. 1.7).[25] Axial charge pairs are formed between a positive amino acid in the Yaa position reaching down to a negatively charged amino acid in the Xaa position of the next triplet in an adjacent chain (Fig. 1.7).[25] These charge pairs are typically have a lysine as the positive amino acid and glutamic acid or aspartic acid as the negatively charged amino acid.[20] Axial interactions have been shown to direct the folding of triple helices, with the ability to make composition and register specific triple helices

as well as staggered triple helices with more than the typical one amino acid offset.[17,19,26,27]

1.4 Creating Composition and Register Specific Triple Helices

The design possibilities for heterotrimeric triple helices are far greater than homotrimeric triple helices, simply because there is a much larger sequence space with two or three distinct chain sequences. One could imagine recreating active sites from natural collagen, creating topologically distinct scaffolds, or unique biomimetic sequences. Heterotrimers offer unique design possibilities, but come with a complex set of design rules that are still not fully understood.

1.4.1 ABC Triple Helices

Heterotrimeric design started with bulk attraction and repulsion of positive and negative charges. A triple helix was designed with ten positively charged resides, $(PRG)_{10}$, a peptide with ten negatively charged residues, $(EOG)_{10}$, and a neutral peptide $(POG)_{10}$.[28] This system showed a distinct triple helix formed only when the three peptides were mixed together, that was not present in a mixture of two of any of the three component peptides. This helix has a melting point of 54 °C while $(POG)_{10}$ had a melting point of 67.5 °C under the same conditions.[29] Expanding on this study, $(DOG)_{10}$, $(EOG)_{10}$, $(PRG)_{10}$, and $(PKG)_{10}$ were all assessed to find the most stable combination of positive and negative amino acids. They found the most stable combination was $(PKG)_{10}$, $(DOG)_{10}$, $(POG)_{10}$ with a melting point of 65 °C which is surprisingly only depressed 2.5 °C from $(POG)_{10}$ when compared to the mixture of just $(PKG)_{10}$ and $(DOG)_{10}$ which melts at 36 °C.[28]

The most stable of these triple helices ($(PKG)_{10}$, $(DOG)_{10}$, $(POG)_{10}$) was examined further with x-ray crystallography and nuclear magnetic resonance spectroscopy (NMR).[30] Through this experiment, it was found that the lysine and aspartate side chains form an axial charge pair across the triple helix, enabling the specific formation of one composition and register in solution. Seeing this as a design tool, axial charge pairs were utilized to computationally design an ABC composition and register specific triple helix.

This use of axial charge pairs to specify composition and register spurned a new area of study where a specific triple helix can be rationally designed. Axial charge pairs have been used to design AAB and ABC composition and register specific triple helices, triple helices that lack hydroxyproline, composition and register specific disease models.[18,19,21,26,31,32]

1.4.2 AAB Triple Helices

The creation of AAB composition and register triple helices also utilized axial charge pairs.[18] The sequence space of AAB composition helices is limited because one chain repeats; typically leaving unpaired charged amino acids, and less sequence space. AAB helices are important because they are the same composition as collagen type I, therefore can be used in disease models.[12]

The AAB composition has been achieved by designing a helix with a net neutral charge, but with individual peptides that have an overall charge. Specifically the system $(PKGPOG)_5$ and $(EOG)_{10}$ mixed in a 2:1 ratio gives a net neutral helix.[33]This triple helix was also observed to have axial charge pairing by NMR. However, two registers of this system where present, $(PKGPOG)_5$ $(EOG)_{10}$ $(PKGPOG)_5$ and $(PKGPOG)_5$ $(PKGPOG)_5$

Figure 1.8. Found registers of the $(PKGPOG)_5$ $(EOG)_{10}$ system. Amino acids in axial charge pairs shown in grey, while unpaired amino acids are shown in yellow.

$(EOG)_{10}$. Both registers have the same amount of charge pairs, and unpaired amino acids (see Fig. 1.8). In order to have register control, this design must be adjusted to stabilize one specific register over the other.[33]

For an AAB composition and register specific sequence, a triple helix was designed where the AAB register was stabilized over the other registers.[18] This was created by taking a homotrimer sequence $(PKGXOG)_5$ (where X is aspartate or glutamate) and removing charged amino acids that would not be charge paired in the triple helix. The resulting helix which consists of two $(PKGXOG)_5$ strands and one $(POGXOG)_5$ strand, alternatively the negatively charged amino acid could be replaced with positively charged amino acids resulting in $(PKGXOG)_5$ and $(PKGPOG)_5$. Of these systems only $(PKGEOG)_5$-$(POGDOG)_5$ and $(PKGDOG)_5$-$(POGDOG)_5$ were found to form composition specific heterotrimers in solution. However, only the $(PKGEOG)_5$-$(POGDOG)_5$ was found to form a single AAB triple helix in solution. $(PKGDOG)_5$-$(POGDOG)_5$ was composition, but not register specific.[18]

1.4.3 Hydroxyproline and Proline Effects

Proline, the only imino acid, and its post-translational modification, hydroxyproline, are ubiquitous in collagen.[7] Their ringed structure is predisposed to the

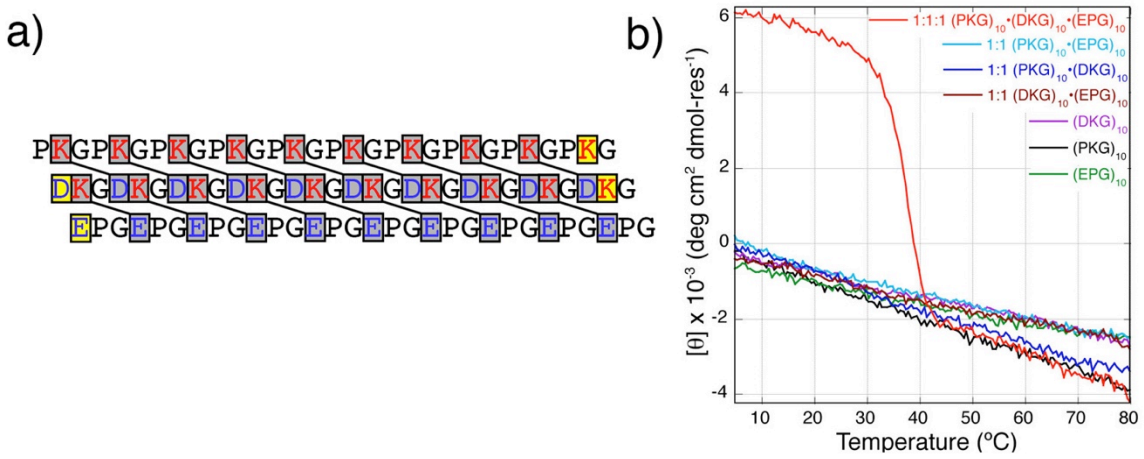


Figure 1.9. (a)Sequence and charge pairs of hydroxyproline-free triple helix. (b)Thermal denaturation monitored by CD of all possible component triple helices. Figure adapted from Jalan et al, 2013.

poly-proline type II secondary structure with phi and psi angles of -75° and 150°, respectively, and they lend stability to the triple helix.[34] However, when designing and producing triple helices they can be problematic. Their presence in a triple helix is stabilizing to all compositions and registers regardless of the sequence unlike charged amino acids that are destabilizing without an oppositely charged partner in the correct position. This creates challenges in complicated composition and register control because the constituent peptides may form homotrimers or undesired heterotrimers if they are stable enough to compete with the target state. Beyond design concerns, a typical method of producing proteins is through bacterial expression. Most bacteria do not posses the enzyme proyl hydroxylase which converts proline to hydroxyproline,[35] and even if they did there is currently no way to signal hydroxylation at specific prolines.

While hydroxyproline is highly stabilizing to a triple helix, the use of axial charge pairs allows a triple helix to be formed without it. This design with eighteen charge pairs and no hydroxyproline melts at 37 °C. It is an ABC composition triple helix of $(PKG)_{10}$, $(DKG)_{10}$, and $(EPG)_{10}$ (Fig. 1.9).[26] None of the alternate compositions form in part

because alternate compositions would have a net charge, with the exception of the homotrimer of $(DKG)_{10}$, which is inherently destabilized with no proline or hydroxyproline. This triple helix is the first collagen mimetic triple helix without hydroxyproline.[26]

1.4.4 Design Principles

Currently, the design of triple helices can be governed by a few principles. First, eliminate unpaired charged amino acids. Having unpaired charged amino acids in the final design decrease the specificity of the target system, not just by destabilizing the target system, but also by potentially participating in charge pairs in other compositions and registers, increasing the stability of those species. Secondly, when creating charge pairs lysine-aspartate pairs are more stabilizing to the triple helix than lysine-glutamate charge pairs. However when designing a composition and register triple helix, it must be kept in mind that all lysine-aspartate charge pairs will be stabilizing to the alternate composition and registers as well. At times a mixture of lysine-aspartate and lysine-glutamate pairs is useful to strike the balance between stability and specificity. The last strategy to increase the specificity of a single composition and register is to change the hydroxyprolines to prolines in the peptide strands that are forming the alternate helices. This will destabilize the alternate helices and the target helix, but if the original helix was stable enough, the correct composition and register will fold.

1.5 Applications of Collagen Mimetic Peptides

Collagen mimetic peptides are a malleable model system for natural collagen. By adjusting the amino acid sequence, triple helices can be made for specific modeling of disease states, for example the disease Osteogenesis Imperfecta,[36] substrates for studying

enzymatic activity[37] or targeted adhesion to damaged collagen,[15] or for the creation of unique supramolecular structures including fibers and nanosheets.[38-41]

1.5.1 Osteogenesis Imperfecta

Osteogenesis Imperfecta is a disease which disrupts the canonical folding of the collagen triple helix,[6,10] specifically the hydrogen bonds that form between the backbone of the different chains. This disease results from mutations to the amino acid sequence of type I collagen, the most harmful of which is to the glycine of the repeating Xaa-Yaa-Gly motif. These mutations result in an array of phenotypes, ranging from easily broken bones to fatality *in utero*.[42,43] This array of phenotypes results from substitutions of diverse amino acids at many possible mutation sites.

From prior studies we know there are a few rules governing the impact of an OI mutation.[44] First, the identity of the substituted amino acid is somewhat predictive of the lethality of the mutation. There are seven most common substitutions alanine, serine, cysteine, valine, aspartate, glutamate, and arginine. The lethality of the mutations is Val >Asp >Glu >Arg >Cys >Ser >Ala for the alpha1 chain and Glu >Asp >Cys >Arg >Val >Ser >Ala for the α2 chain,[42] which is the same as the trend seen in model systems.[45] In systems where there is a possibility of the substituted side chain hydrogen bonding seen in serine mutations, there can be a stabilizing effect of the side chain hydrogen bonding to the backbone of the other chains.[46,47] Secondly, the mutation could appear in both α1 chains or just the α2 chain, where the two mutations have a greater degree of disruption than one.[42,48] Third, the rigidity of the surrounding amino acids affects the ability of the triple helix to renucleate after disruption.[49,50]

Lastly, the location of the mutations within collagen sequence can have an effect. If the mutation occurs close to a binding region the mutation would be more lethal, and the mutations close to the N-terminus are less impactful than the mutations close to the C-terminus.[42,43,51]

Previous model systems have elucidated the loss of stabilizing hydrogen bonds in the mutated triple helix. A $(POG)_4POA(POG)_5$ homotrimeric crystal structure has shown four backbone hydrogen bonds lost.[52] Subsequent NMR studies have shown two hydrogen bonds lost in the solution state.[46] These systems were both homotrimeric triple helices, which do not directly reflect the natural AAB composition of type I collagen.[53]

The first study, which displayed a mutation in a heterotrimeric triple helix, was in a ABC composition system.[54] A guest region was incorporated with an AAB composition and that region contained mutations in one, two, or three chains. Unfortunately that study only included Circular Dichroism (CD) analysis, which cannot give register information. However, this study was able to show that the first mutation to the triple helix was the most impactful, with subsequent mutations having a smaller effect.

Other collagen types have been studied, for example type IV with interruptions to the canonical X-Y-Gly triplet. It was a heterotrimeric system with disruptions in both chains, namely a glycine deletion in the A chain and a glycine to leucine mutation in the B chain. This study surprisingly saw controlled composition and register without charge pairs, and the concluded that the mutations had an impact on the register and composition.[55]

1.5.2 Fiber Forming Peptides

The natural collagen fiber is a complex supramolecular assembly with forces elongating the triple helix as well as forces assembling the triple helices side-by-side resulting in a wide and long fiber. Natural fiber formation is controlled by globular domains at the ends of the collagen protein strands which are later cleaved.[12] However, in collagen mimetic peptides globular domains would be too large and dominate the secondary structure, so other assembling forces must be used. Commonly these fibers are assembled through the used of charge pairing, however a group has reported fiber formation with the use of pi-pi stacking interactions.[40]

Beyond fiber formation, a true collagen mimic would have aligned and organized fibers. In natural collagen this alignment is evident in the d-periodicity of the fibrils. Only one group has succeeded in creating d-periodic collagen mimetic fibers, which they were able to grow to a length of 3-4 micrometers and a width of 12-15 nanometers. However these fibers, were only d-periodic in a specific buffer system, and had a low melting point between 37 and 48 °C depending on the buffer system.[38]

Other fiber forming peptides result in hydrogel formation, which could have a wide range of translational applications as scaffolds for bone formation, hemostats, and wound healing. These hydrogels are formed utilizing axial charge pairs, which stabilize and direct the assembly.[56] In some cases they are also hypothesized to offset the peptide strands, resulting a sticky-ended intermediate that catalyzes fiber formation.[39,41]

1.5.3 Alternate Supramolecular Structures

Beyond fibers, other supramolecular structures, specifically sheets, have been formed by modification of the collagen mimetic peptide sequence. These sheets can be formed by peptides with positive, negative, and neutral regions that form triple helices

and pack in an antiparallel fashion, matching a positive region to a negative region.[57] In the most controlled and homogenous sheets, the positive region utilized aminoproline (both the 4R and 4S version), an unnatural amino acid. Sheets can also be formed by a mixture of positively charged triple helices and negatively charged triple helices leading to matching oppositely charge tips aligning in an upright fashion, as hypothesized by the researchers.[58] A two-peptide composite has also been used to form multilayered sheets of positive and negative triple helices. Both of these sheet structures needed arginine to mediate the helix-helix interactions, and there was no supramolecular structure formed when lysine was used.[59,60]

1.5.4 Biological Applications

Collagen mimetic peptides have been used extensively to study interaction with other proteins and enzymes. Toolkits have been designed from the sequences of type I, II, and III collagen where the sequences are broken up into twenty-seven amino acid chunks with overlapping regions on the tips.[61] These toolkits have been used to interrogate binding sites in collagen with a range of proteins such as integrin α2β1, platelet receptor Gp VI, plasma protein vWF, a tyrosine kinase receptor DDR2, and basement membrane protein BM-40. These studies elucidate binding sites like GXYGER motifs with integrin α2β1.[61]

Collagen mimetic peptides have also been used as a model system to study the activity of matrix metalloproteases (MMP),[37] which cleave proteins including collagens. Short collagen mimetic peptides with native collagen sequences were used to elucidate the cleavage site of hundreds of MMPs. These studies were also expanded to examine the relative rates of MMP activity against the mimetic peptides.[62]

1.6 Conclusions and Preview of Thesis

The collagen mimetic peptide is a versatile model system and building block for supramolecular assembly. This class of peptides has a well-understood synthesis and the design principles, while not complete, can be used to design specific and reproducible triple helices. With improvements in the stability of the heterotrimeric triple helix, these peptides have the potential to be an important model system in understanding collagen's role in biological processes.

In this thesis I will present improvements in the collagen mimetic peptide model system. In Chapter 2, I will discuss innovations in the design of collagen mimetic peptides. Axial charge pairs were used to create longer offsets in a triple helix, increasing our understanding of the stability conferred by charge paring. This study differentiates the stability of axial and lateral charge pairs, where axial charge pairs are found to be more stabilizing than their lateral counterpart. In this chapter, experiments were designed by Dr. Jalan and Prof. Hartgerink, with peptides synthesized by Dr. Jalan and I. I performed purification and CD studies, as well as a portion of the NMR spectra collection, with the rest completed by Dr. Jalan. Dr. Jalan, Prof Hartgerink, and I analyzed the data and prepared the manuscript.

In Chapter 3, I will discuss the development of a self-assembling AAB composition and register controlled model of the disease Osteogenesis Imperfecta. Here I will focus on mutations to the B chain of the AAB triple helix. This model is the first collagen mimetic triple helix that incorporates a mutation like those seen in Osteogenesis Imperfecta while folding into a designed composition and register. In chapters 3 and 4 the studies were conceived and designed by Amanda Acevedo-Jake, Prof. Hartgerink, and

myself. Amanda Acevedo-Jake and I split the synthesis and purification of the peptides, while I completed CD analysis. Amanda Acevedo-Jake collected and analyzed the NMR spectra, while I translated the spectra to models using Rosetta. Amanda Acevedo-Jake, Prof. Hartgerink, and myself analyzed the data and prepared the manuscripts. In Chapter 4, Douglas Walker created the algorithm to calculate helical twist and assisted in analysis of that data.

In Chapter 4, I will expand on the model developed in Chapter 3 with mutations to the A chain of the AAB triple helix. These mutations resulted in disruptions to the composition and register of the model, the first time composition disruption has been reported as the result of mutations to glycines in the triple helix. This disruption was explored further with mutations in both the A and B chains of the AAB triple helix, which was found to fold with controlled composition and register. This study also developed a new parameter to analyze the triple helical topology called helical twist. These studies have contributed a new understanding to the forces stabilizing and destabilizing the triple helix.

1.7 References

(1) Chattopadhyay, S.; Raines, R. T. *Biopolymers* **2014**, *101*, 821-833

(2) Lee, K. Y.; Mooney, D. J. *Chem. Rev.* **2001**, *101*, 1869-1879.

(3) Friess, W. *Eur. J. Pharm. Biopharm.* **1998**, *45*, 113-136.

(4) Delgado, L. M.; Bayon, Y.; Pandit, A.; Zeugolis, D. I. *Tissue Eng., Part B* **2015**, *21*, 298-313.

(5) Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.

(6) Ramachandran, G.; Sasisekharan, V. *Nature* **1954**, *190*, 1004-1005.

(7) Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry*

2000, *39*, 14960-14967.

(8) Chan, V. C.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B.; Beck, K. *J. Biol. Chem.* **1997**, *272*, 31441-31446.

(9) Ramshaw, J. A. J.; Shah, N. K. N.; Brodsky, B. B. *J. Struct. Biol.* **1998**, *122*, 86-91.

(10) Rich, A.; Crick, F. H. C. *J. Mol. Biol.* **1961**, *3*, 483-506.

(11) Harrington, W. F. *J. Mol. Biol.* **1964**, *9*, 613-617.

(12) Ricard-Blum, S. *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, 1-19.

(13) Baum, J.; Brodsky, B. *Curr. Opin. Struct. Biol.* **1999**, *9*, 122-128.

(14) Hodges, J. A.; Raines, R.T. *J. Am. Chem. Soc.* **2003**, *125*, 9262-9263.

(15) Li, Y.; Yu, S. M. *Curr. Opin. Chem. Biol.* **2013**, *17*, 1-8.

(16) Buevich, A.; Baum, J. *Philos. Trans. R. Soc., B* **2001**, *356*, 159-168.

(17) Fallas, J. A.; Hartgerink, J. D. *Nat. Commun.* **2012**, *3*, 1-8.

(18) Jalan, A. A.; Hartgerink, J. D. *Biomacromolecules* **2013**, *14*, 179-185.

(19) Jalan, A. A.; Jochim, K. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2014**, *136*, 7535-7538.

(20) Fallas, J. A.; O'Leary, L. E. R.; Hartgerink, J. D. *Chem. Soc. Rev.* **2010**, *39*, 3510-3527.

(21) Fallas, J. A.; Lee, M. A.; Jalan, A. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2012**, *134*, 1430-1433.

(22) Wei, F.; Fallas, J. A.; Hartgerink, J. D. *Macromol. Rapid Commun.* **2012**, *33*, 1445-1452.

(23) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2005**, *44*, 1414-1422.

(24) Stultz, C. M. *Protein Sci.* **2006**, *15*, 2166-2177.

(25) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(26) Jalan, A. A.; Demeler, B.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2013**, *135*, 6014-6017.

(27) Jalan, A. A.; Hartgerink, J. D. *Curr. Opin. Chem. Biol.* **2013**, *17*, 960-967.

(28) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034-15041.

(29) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683-2690.

(30) Fallas, J. A.; Dong, J.; Tao, Y. J.; Hartgerink, J. D. *J. Biol. Chem.* **2012**, *287*, 8039-8047.

(31) Clements, K. A.; Acevedo-Jake, A. M.; Walker, D. R.; Hartgerink, J. D. *Biomacromolecules* **2017**, *18*, 617-624.

(32) Acevedo-Jake, A. M.; Clements, K. A.; Hartgerink, J. D. *Biomacromolecules* **2016**, *17*, 914-921.

(33) O'Leary, L. E. R.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2011**, *133*, 5432-5443.

(34) Adzhubei, A. A.; Sternberg, M. J. E.; Makarov, A. A. *J. Mol. Biol.* **2013**, *425*, 2100-2132.

(35) Mohs, A. A.; Silva, T. T.; Yoshida, T. T.; Amin, R. R.; Lukomski, S. S.; Inouye, M. M.; Brodsky, B. B. *J. Biol. Chem.* **2007**, *282*, 29757-29765.

(36) Bhate, M.; Wang, X.; Baum, J.; Brodsky, B. *Biochemistry* **2002**, *41*, 6539-6547.

(37) Nagase, H.; Fields, G. B. *Biopolymers* **1996**, *40*, 399-416.

(38) Rele, S.; Song, Y.; Apkarian, R. P.; Qu, Z.; Conticello, V. P.; Chaikof, E. L. *J. Am. Chem. Soc.* **2007**, *129*, 14780-14787.

(39) Xu, F.; Silva, T.; Joshi, M.; Zahid, S.; Nanda, V. *J. Biol. Chem.* **2013**, *288*, 31616-31623.

(40) Cejas, M. A.; Kinney, W. A.; Chen, C.; Leo, G. C.; Tounge, B. A.; Vinter, J. G.; Joshi, P. P.; Maryanoff, B. E. *J. Am. Chem. Soc.* **2007**, *129*, 2202-2203.

(41) O'Leary, L. E. R.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, J. D. *Nat. Chem.* **2011**, *3*, 821-828.

(42) Marini, J. C.; Forlino, A.; Cabral, W. A.; Barnes, A. M.; San Antonio, J. D.; Milgrom, S.; Hyland, J. C.; Korkko, J.; Prockop, D. J.; De Paepe, A.; Coucke, P.; Symoens, S.; Glorieux, F. H.; Roughley, P. J.; Lund, A. M.; Kuurila-Svahn, K.; Hartikka, H.; Cohn, D. H.; Krakow, D.; Mottes, M.; Schwarze, U.; Chen, D.; Yang, K.; Kuslich, C.; Troendle, J.; Dalgleish, R.; Byers, P. H. *Hum. Mutat.* **2007**, *28*, 209-221.

(43) Byers, P. H. *Philos. Trans. R. Soc., B* **2001**, *356*, 151-158.

(44) Brodsky, B.; Thiagarajan, G.; Madhan, B.; Kar, K. *Biopolymers* **2008**, *89*, 345-353.

(45) Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. *Proc. Nat. Acad. Sci. U.S.A.* **2000**, *97*, 4273-4278.

(46) Li, Y.; Brodsky, B.; Baum, J. *J. Biol. Chem.* **2009**, *284*, 20660-20667.

(47) Mooney, S. D.; Klein, T.E. *Mol. Cell. Proteomics* **2002**, *1*, 868-875.

(48) Byers, P. H.; Cole, W. G. Osteogenesis Imperfecta. In In *Connective Tissue and Its Heritable Disorders*; Royce PM, Steinmann B, Eds.: John Wiley &Sons Inc.: Hoboken, NJ, 2003; pp. 385–430.

(49) Yang, W.; Battineni, M. L.; Brodsky, B. *Biochemistry.* **1997**, *36*, 6930-6935.

(50) Radmer, R. J.; Klein, T. E. *Biochemistry* **2004**, *43*, 5314-5323.

(51) Hyde, T. J.; Bryan, M. A.; Brodsky, B.; Baum, J. *J. Biol. Chem.* **2006**, *281*, 36937-36943.

(52) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. E. *Science.* **1994**, *266*, 75-81.

(53) Kielty, C. M.; Hopkinson, I.; Grant, M. E. Collagen: The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix. In *Connective Tissue and Its Heritable Disorders*; Royce, P. M., Steinmann, B., Eds.; John Wiley & Sons Inc.: Hoboken, NJ, 2003; pp.103-147.

(54) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7509-7515.

(55) Xiao, J.; Sun, X.; Balaram, M.; Brodsky, B.; Baum, J. *J. Biol. Chem.* **2015**, *192*, 127-137.

(56) Xu, F.; Li, J.; Vikas, J.; Tu, R. S.; Huang, Q.; Nanda, V. *J. Am. Chem. Soc.* **2011**, *134*, 47-50

(57) Jiang, T.; Xu, C.; Liu, Y.; Liu, Z.; Wall, J. S.; Zuo, X.; Lian, T.; Salaita, K.; Ni, C.; Pochan, D.; Conticello, V. P. *J. Am. Chem. Soc.* **2014**, *136*, 4300-4308.

(58) Jiang, T.; Vail, O. A.; Jiang, Z.; Zuo, X.; Conticello, V. P. *J. Am. Chem. Soc.* **2015**, *137* , 7793-7802,

(59) Parmar, A. S.; James, J. K.; Grisham, D. R.; Pike, D. H.; Nanda, V. *J. Am. Chem. Soc.* **2016**, *138*, 4362-4367.

(60) Jiang, T.; Xu, C.; Zuo, X.; Conticello, V. P. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 8367-8371.

(61) Farndale, R. W.; Lisman, T.; Bihan, D.; Hamaia, S.; Smerling, C. S.; Pugh, N.;

Konitsiotis, A.; Leitinger, B.; de Groot, P. G.; Jarvis, G. E.; Raynal, N. *Biochem. Soc. Trans.* **2008**, *36*, 241-250.

(62) Lauer-Fields, J. L.; Tuzinski, K. A.; Shimokawa, K.-I.; Nagase, H.; Fields, G. B. *J. Biol. Chem.* **2000**, *275*, 13282-13290.

Chapter 2: Rational Design of a Non-canonical "Sticky-Ended" Collagen Triple Helix*

2.1 Introduction

In DNA origami,[1] the adenine−thymine and guanine−cytosine nucleotide pairing is used to create sticky-ended motifs that self-assemble into complex nanostructures.[2-4] Similarly, in α-helical coiled-coil design, interactions between side chains of acidic and basic amino acids patterned at the e and g positions of the abcdefg heptad repeat introduce offset in an otherwise blunt-ended oligomer and nucleate fiber formation.[7,8] Therefore, design of protein folds containing unsatisfied polar or electrostatic contacts at the N- or C-termini is of interest from a supramolecular assembly perspective. Sticky-ended motifs can be routinely designed for DNA and α-helical coiled coils. Although a covalently linked sticky-ended collagen triple helix has been reported before,[9] proof-of-principle design of a non-covalently self-assembled sticky-ended collagen triple helix has remained elusive. Here we use axial salt-bridge interactions to place the three peptides of a collagen triple helix in a non-canonical four-amino-acid offset. The sticky ends are short and do not promote fiber formation, which allows us to confirm the intended molecular arrangement within the triple helix.

In a collagen triple helix, three peptides containing Xaa-Yaa-Gly triplet amino acid repeats supercoil, with a one-amino-acid offset between the peptide chains. The one-

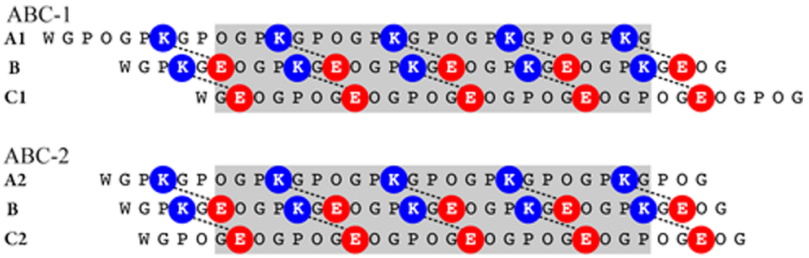
---

* This chapter was adapted from Jalan, A.A; Jochim, K.A.; Hartgerink, J.D.; "Rational Design of a Non-Canonical "Sticky-Ended" Collagen Triple Helix" *J. Am. Chem. Soc*. **2014**, 136, 7535-7538.

This paper was published under the thesis author's maiden name of Jochim.

KAJ and AAJ synthesized, purified, and collected NMR data. KAJ performed CD analysis. KAJ, AAJ, and JDH analyzed the data and wrote the manuscript.

residue offset maximizes inter-peptide hydrogen bonds and optimizes helical packing.[10] Gly residues form the helical core, and therefore higher order assembly of the triple helices can only be mediated by patterning residues in the solvent-exposed Xaa and Yaa positions.[11] Using short peptides that mimic the triple-helical protein fold of native collagen, we recently elucidated sequence rules for interpeptide salt-bridge interactions between a Yaa- positioned Lys and an Xaa-positioned Asp or Glu residue.[12] We call these "axial" salt-bridges, as both side-chain conformations are approximately coaxial with the helical axis. Each Gly residue of the triple helix defines a cross-sectional plane that also contains the Xaa and Yaa residues of the other two chains. In terms of sequence position, an axial salt-bridge forms between the Lys in the nth and the Asp or Glu in the n+3 cross-sectional plane. The energetic stabilization derived from the interaction between Lys and Asp/Glu residues present within the same cross-sectional plane, which we call a "lateral" interaction, has been shown to be conformationally less robust than an axial salt-bridge.[13]



Scheme 2.1. Sequences of non-canonical offset ABC-1 and canonical offset ABC-2 heterotrimers, indicating positions of Lys–Glu axial salt-bridges. Grey highlighted area indicates the core region, which is identical in both triple helices.

We have used axial salt-bridges to design AAA-,[14] AAB-[15,16] and ABC-type [17-19] collagen triple helices. Of particular interest here is the peptide $(PKG)_4(POG)_4(DOG)_4$ that rapidly self- assembles into long fibers and ultimately a hydrogel in aqueous solution.[14] To explain the rapid self-assembly, it was proposed that the peptide forms a

transient homotrimeric collagen triple helix with a 20-amino-acid offset between the peptide chains. The unusually long offset would contain numerous unsatisfied hydrogen-bonding partners as well as unpaired and electrostatically charged Lys and Asp residues, which could drive rapid aggregation. However, due to the fibrous morphology of the aggregates, the offset within the collagen triple helix could not be elucidated using any biophysical tool available. Control over peptide offset in a triple helix has the potential to provide unprecedented control over its supramolecular fate. Therefore, in this chapter we demonstrate a proof-of-principle design of a collagen triple helix, ABC-1, with a four-residue offset between the peptide chains (Scheme 2.1). The short offset reduces the likelihood of aggregation, allowing us to study its structure and stability in molecular detail by NMR. A second collagen triple helix, ABC-2, with identical amino acid composition but containing a one-amino-acid offset, is also designed. These two systems demonstrate our ability to modulate chain offsets in a collagen triple helix and lay the groundwork for design of a new class of collagen-related nanostructures.

2.2 NMR and CD Analysis of ABC-1 and ABC-2

Assuming a canonical single-amino-acid offset between peptides, a mixture of three peptides, A, B, and C, can self- assemble into 27 unique triple helices. Axial Lys−Asp/Glu salt- bridges have been used to bias the ensemble of competing states to selectively populate one AAB-[16] or ABC-type[17] heterotrimer of desired chain composition and register. In the case of ABC-1, we placed Lys and Glu residues such that they only form axial salt-bridges if the peptide chains adopt a four- residue offset, therefore setting the stage for an offset or sticky- ended assembly.

The thermal melting curve of ABC-1 shows a broad transition with a major

Sequence	Abbreviation	T_m °C
WG(POGPKG)$_2$PO**G**PKG(POGPKG)$_2$	A1	-
WG(PKGPOG)$_2$PKGPO**G**(PKGPOG)$_2$	A2	-
WG(PKGEOG)$_2$PK**G**EOG(PKGEOG)$_2$	B	32
WG(EOGPOG)$_2$EO**G**POG(EOGPOG)$_2$	C1	45
WG(POGEOG)$_2$PO**G**EOG(POGEOG)$_2$	C2	42
WG(POGPKG)$_5$/ WG(PKGEOG)$_5$	A1B	36
WG(PKGPOG)$_5$/ WG(PKGEOG)$_5$	A2B	39
WG(PKGEOG)$_5$/ WG(EOGPOG)$_5$	BC1	44
WG(PKGEOG)$_5$/ WG(POGEOG)$_5$	BC2	41
WG(POGPKG)$_5$/ WG(EOGPOG)$_5$	A1C1	m
WG(POGPKG)$_5$/ WG(POGEOG)$_5$	A2C2	m
WG(POGPKG)$_5$/WG(PKGEOG)$_5$/ WG(EOGPOG)$_5$	ABC-1	47
WG(POGPKG)$_5$/WG(PKGEOG)$_5$/ WG(POGEOG)$_5$	ABC-2	52

Table 2.1. Sequences, Abbreviations, and Melting Temperatures of Triple Helices Studied. Bold glycines indicate ^{15}N isotopically labeled amino acids. M indicates a broad melting transistion.

unfolding event at 47 °C (Figure 2.1a). In addition to a substantial decrease in van der Waals surface area, ABC-1 loses four inter-peptide hydrogen bonds at the termini due to the four-residue offset. The resulting conformational disorder is expected to result in a broader thermal transition, as terminal regions unfold before the more stable core of the triple helix. In comparison, ABC-2 shows a cleaner thermal transition, with a higher T_m value of 52 °C (Table 2.1).

Due to the difference in hydrogen bonds formed by the ^{15}N- Gly amide protons, each of the competing triple-helical states in a mixture of three peptides yields a fingerprint spectrum in a ^{1}H,^{15}N heteronuclear single-quantum coherence (HSQC) experiment. The HSQC spectra of ABC-1 shows three trimer cross-peaks labeled T_1, T_2, and T_3 (Figure 2.1d), which do not overlap with cross-peaks of individual peptides or their binary mixtures. This suggests that mixing peptides A1, B, and C1 results in only one triple helix of A1·B·C1 composition. Additionally, the presence of only three trimer cross-peaks suggests that the three peptides in ABC-1 are arranged in only one register. Similar conclusions can be drawn from the HSQC spectra of ABC-2 and its component

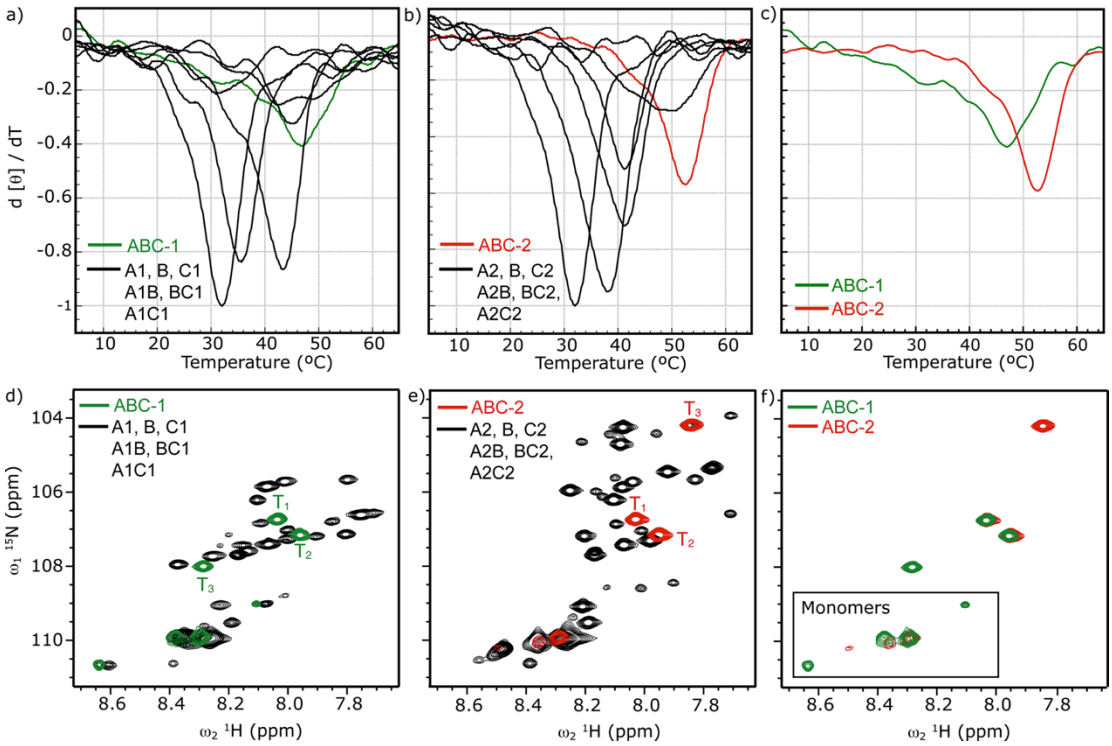


Figure 2.1. (Top) First-derivative plots of thermal unfolding curves of (a) ABC-1 and (b) ABC-2 overlaid with those of component peptides (shown in black) and (c) overlay of ABC-1 and ABC-2 spectra. (Bottom) ^{1}H,^{15}N-HSQC NMR spectra of (d) ABC-1 and (e) ABC-2 overlaid with those of component peptides (shown in black) and (f) overlay of ABC-1 and ABC-2 HSQC spectra, demonstrating the similarity of two of the three N–H peaks. Monomer peaks are indicated within the box.

peptides in Figure. 2.1e. As shown in Figure 1f, two trimer cross-peaks in the HSQC spectra of ABC-1 and ABC-2 have very similar chemical shifts. A comparison of their amino acid sequence in Figure 2.3 indicates that two of the three inter-peptide hydrogen bonds formed by ^{15}N- Gly(NH) are between identical sets of amino acid triplets. In contrast the third hydrogen bond is formed between dissimilar triplets, which explains the observed overlap in only two of the three cross-peaks. Importantly, if the ABC-1 were to adopt a canonical one-residue offset, all three hydrogen bonds, and therefore HSQC peaks, would be chemically distinct from those of ABC-2, and no cross-peak overlap would be expected. This result provides the first molecular-level evidence for the four-residue chain offset in ABC-1.

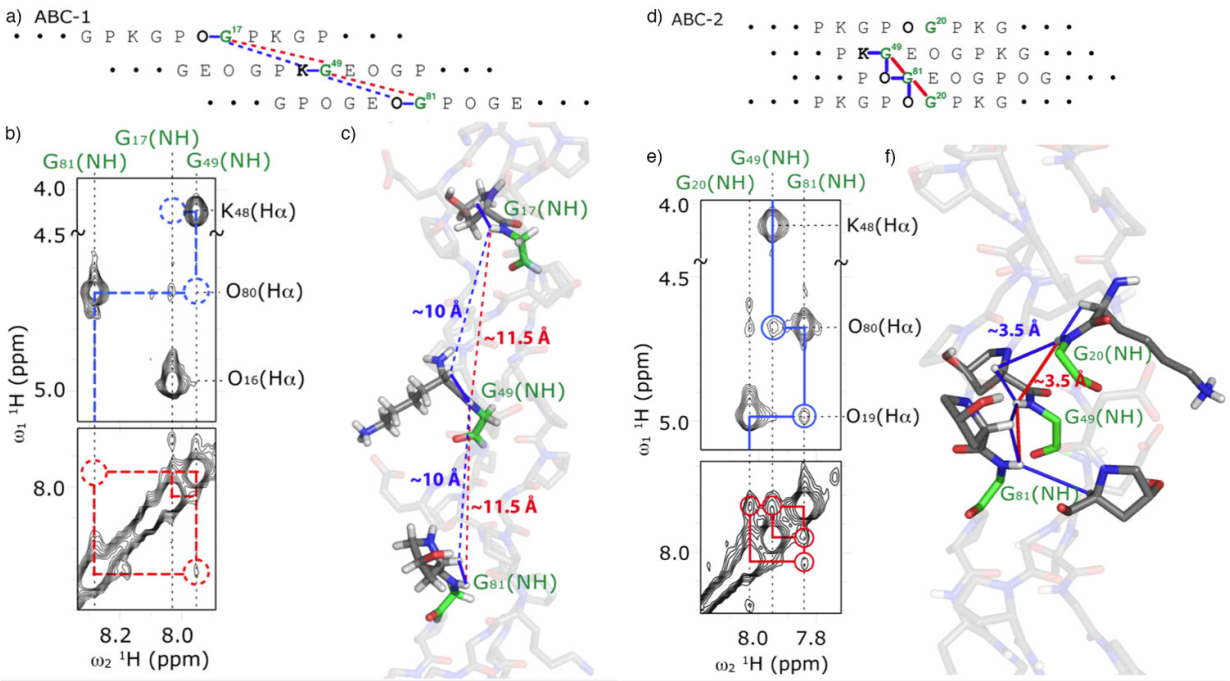


Figure 2.2. (a) Sequence, (b) ^{15}N-edited-NOESY spectrum, and (c) model of ABC-1 indicating absence of inter-peptide ^{15}NH–^{15}NH (red lines) and ^{15}NH–Hα (blue lines) NOE cross-peaks due to >6 Å separation between the protons. (d) Sequence, (e) ^{15}N-edited-NOESY spectrum, and (f) model of ABC-2 indicating absence of inter-peptide ^{15}NH–^{15}NH (red lines) and ^{15}NH–Hα (blue lines) NOE cross-peaks due to >6 Å separation between the protons.

Additional evidence for the intended chain offset comes from homonuclear correlation NMR experiments. A two-dimensional ^{1}H–^{1}H plane of the three-dimensional ^{1}H,^{15}N-NOESY- HSQC experiment (hereafter called edited-NOESY) shows nuclear Overhauser effect (NOE)-based cross-peaks between ^{15}N-Gly(NH) and protons within ~6 Å of it. Since the intensity of the cross-peaks varies as $1/r^6$, where r is the spatial distance between the correlating protons, generally no cross-peaks are observed at distances greater than 6 Å. Due to the four-residue offset in ABC-1, the ^{15}N-Gly amide protons of the A1, B, and C1 chains are more than 11.5 Å apart (Figure 2.2a). As a result, the edited-NOESY spectrum of ABC-1 shows no cross-peaks for correlations between the amide protons (Figure 2.2b, red dotted circles). Additionally, ^{15}N-Gly amide protons are also placed more than 10 Å apart from Hα of the residues preceding ^{15}N-Gly of another chain.

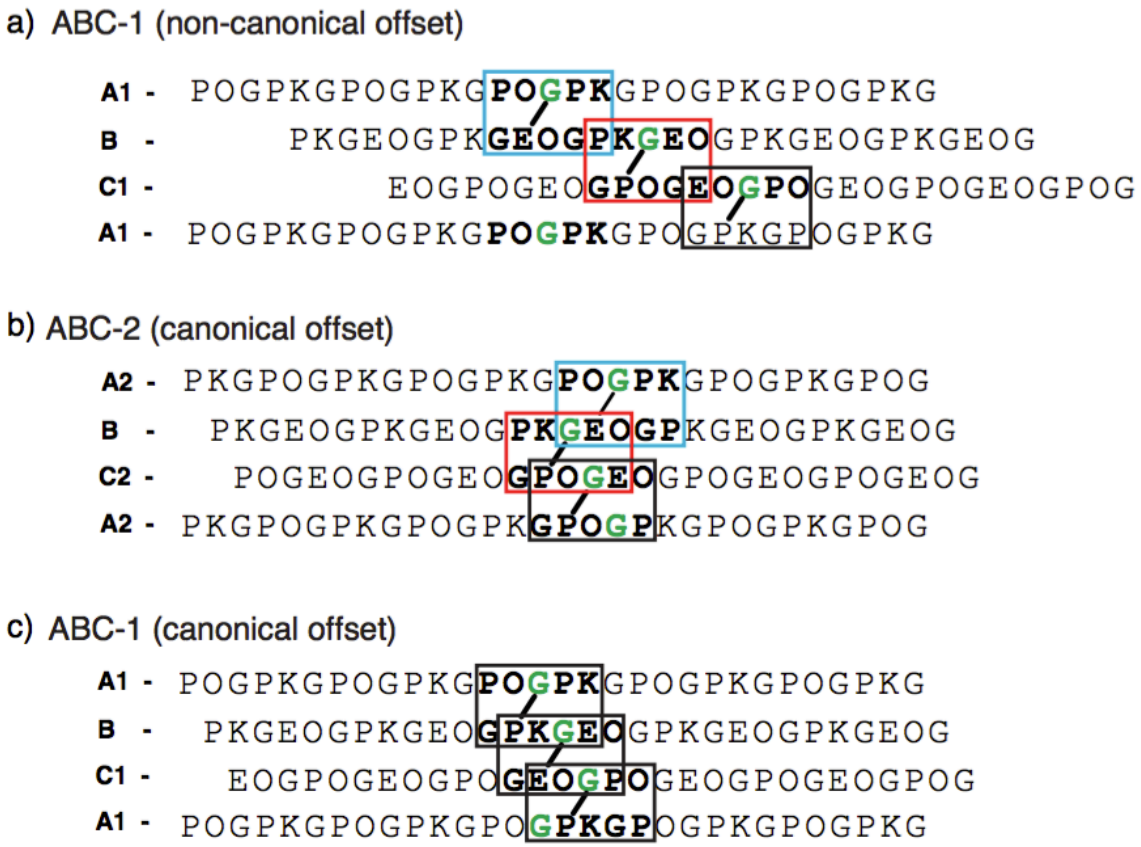


Figure 2.3. Comparison of the hydrogen bonds formed by the ^{15}N-Gly residues (shown in green) in ABC-1 in a non-canonical (a) and ABC-2 in a canonical (b) offset. Peptides in ABC-1 arranged in a hypothetical canonical offset are shown in (c) for comparison. Cyan and red boxes indicate identical chemical environment of ^{15}N-Gly in chains A1 and B of ABC-1 and chains A2 and B of ABC-2. ^{15}N-Gly residues in dissimilar chemical environment are shown within black boxes.

As expected, these ^{15}NH-Hα correlations are also not observed (Figure 2.2b, blue dotted circles).

Analysis of the edited-NOESY spectra of ABC-2 confirms that the absence of NOE cross-peaks in ABC-1 is not an experimental artifact. As shown in Figure 2.2d, the N-^{15}Gly's of A2, B, and C2 chains of ABC-2 heterotrimer are within the observable NOE distance of 6 Å. Therefore, both ^{15}NH–^{15}NH and ^{15}NH–Hα correlations are observed in its edited-NOESY spectra (Figure 2.2e). The accompanying homology model in Figure 2.2f, built from PDB 2KLW,[12] indicates the observed NOEs.

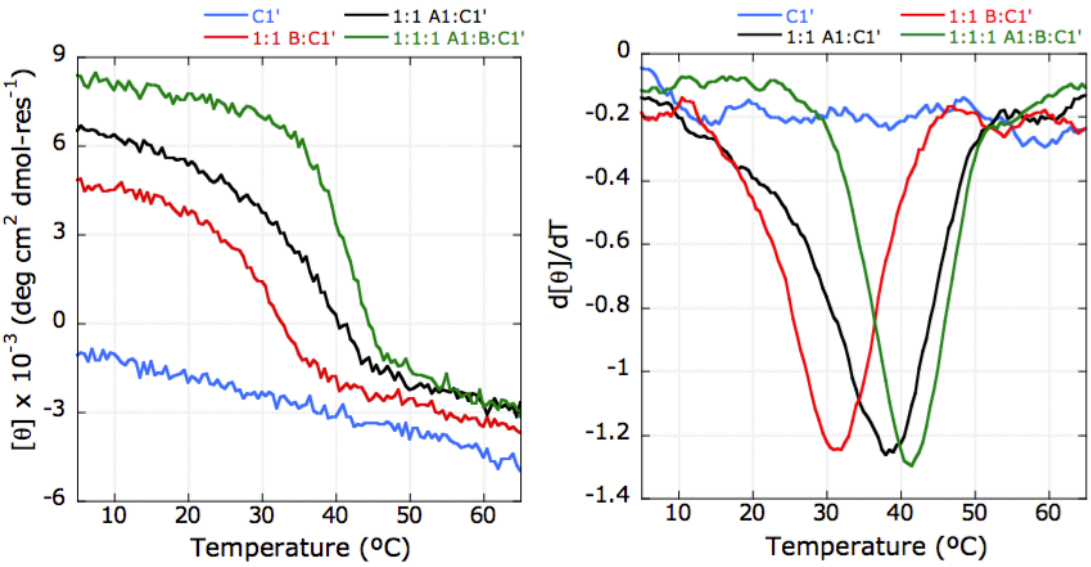


Figure 2.4. CD thermal melting profile of individual, binary and ternary peptide mixtures used for testing A1, B and C1 .

The peptides A2, B, and C2 can arrange in six different ways within the ABC-2 heterotrimer, but pairing of all Lys and Glu residues is only possible in one of them. Comparison of the observed ^{15}NH–^{15}NH and ^{15}NH–Hα correlations to those expected in the six possible peptide arrangements in Table 2.2 confirms that the register of A2, B, and C2 shown in Scheme 2.1 is correct, allowing all Lys and Glu residues to form salt-bridges.

Similarly, the relative chain arrangement in ABC-1 depicted in Scheme 2.1 is critical for salt-bridge formation. To determine the peptide positions in ABC-1, we compared its two- dimensional ^{1}H,^{1}H-NOESY spectra to that of ABC-2. Due to identical triple-helical core regions, their NOESY spectra are nearly identical, indicating similarity in chain arrangement within the triple helix. Furthermore, a NOE cross-peak between E(NH) and K(Hε), characteristic of an axial salt-bridge,[13,19] is observed in the NOESY spectra of both heterotrimers. These results strongly suggest that the relative positions of peptide chains in both heterotrimers are as shown in Scheme 2.1.

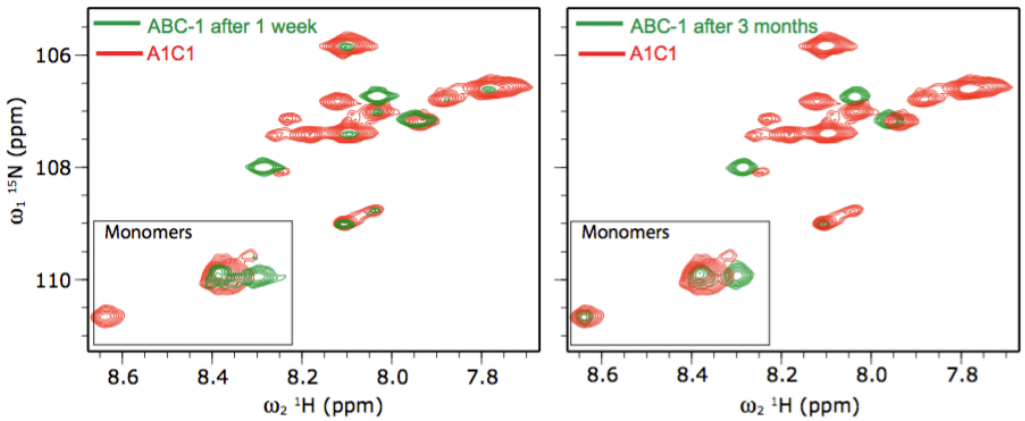


Figure 2.5. ^{1}H,^{15}N-HSQC spectra of ABC-1 one week (left) and three month (right) after mixing A1, B and C1 peptides. The competing heterotrimeric peptide mixture A1C1 is shown in red. The monomer cross peaks are shown in the box.

2.2.1 Direct Evidence of Offset

In order to conclusively demonstrate the sticky- ended four residue chain stagger in ABC-1, we designed a related system ABC-1'. In this heterotrimer, chain C2 with the WG(EOGPOG)$_5$ sequence is replaced with C2' with sequence WG(EPGPPG)$_5$. ABC-1takes nearly three months to achieve single composition and single register and before this time the principle competing state is a binary mixture of A1 and C1. A comparison of the various AAB-type heterotrimers possible in a mixture of A1 and C1 revealed that the AAB composition containing two WG(EOGPOG)$_5$ chains would the most stable. Therefore, we reasoned that a Hyp to Pro substitution in WG(EOGPOG)$_5$ chain should destabilize this competing state and enhance the kinetics of ABC-1' formation. Fig. 2.4 and 2.5 below shows the results of HSQC and CD experiments. The HSQC shown in Fig. 2.4 was acquired two weeks after annealing and shows ABC-1' to be major species with a small population of competing triple helices. In addition to the substitution of Hyp for Pro, we also moved the ^{15}N-Gly to a new position indicated in Fig. 2.6. The new position allows us to observe at least one set of NOE between the ^{15}N- Gly of the middle chain and Hα of the lagging chain which confirms the four residue offset within the ABC-1'

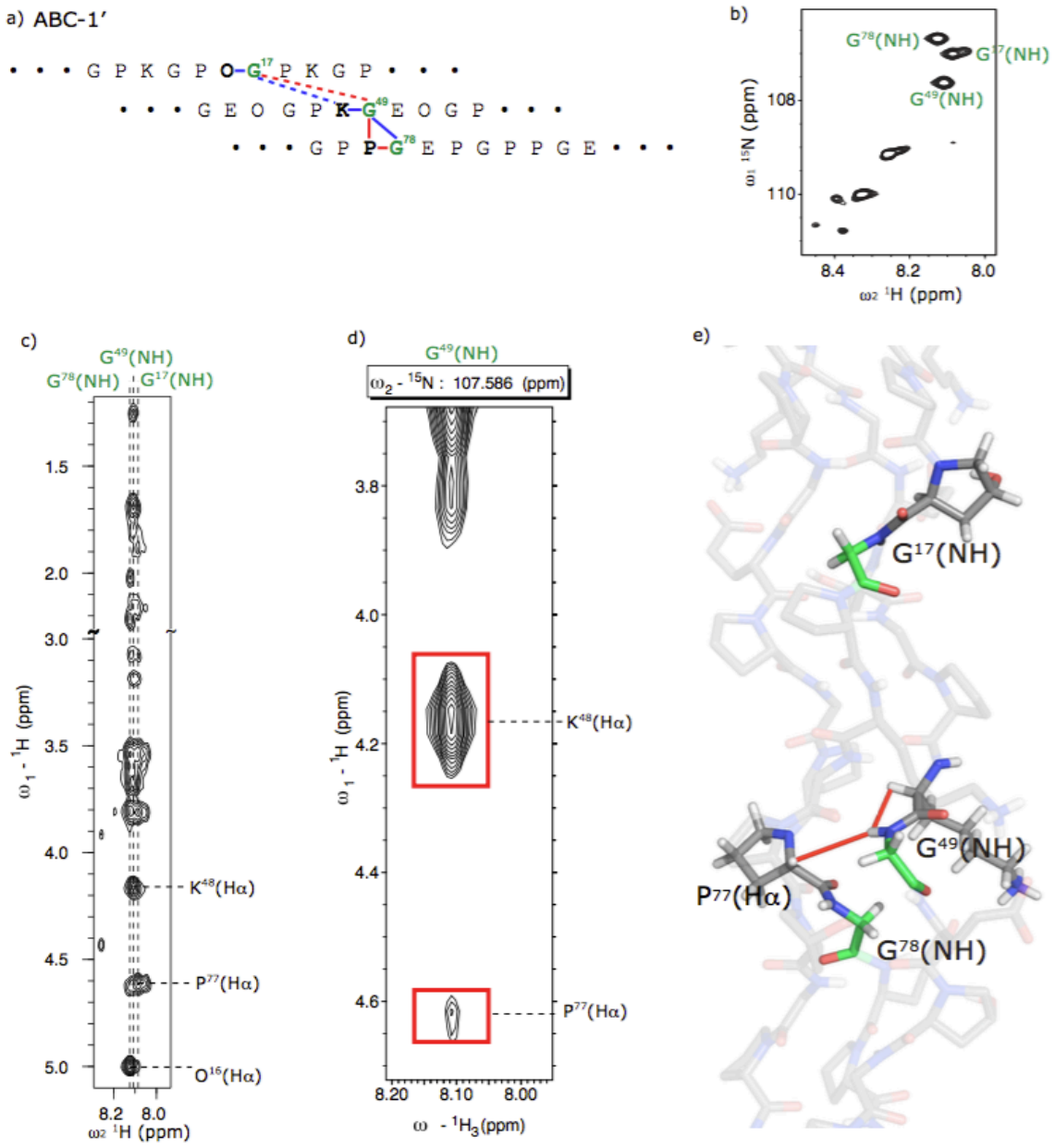


Figure 2.6. a) Position of ^{15}N-Gly labels in ABC-1' shown in green, b) HSQC spectra of ABC-1' acquired on a higher field 750 MHz Bruker cold probe to improve chemical shift dispersion, c) two dimensional ^{1}H-^{1}H version of the 3D-edited-NOESY indicating significant overlap in the amide proton chemical shifts of the three ^{15}N-Gly, d) ^{1}H-^{1}H plane of a three dimensional edited-NOESY experiment at the ^{15}N-chemical shift corresponding to the ^{15}N-Gly of the middle chain. NOEs observed are indicated in model (e).

Observed NOEs	A1·B·C1	B·C1·A1	A1·C1·B	B·A1·C1	C1·A1·B	C1·B·A1
G_{20}(NH)-G_{81}(NH)	✔	✔	✔	✖	✖	✔
G_{81}(NH)-G_{49}(NH)	✔	✖	✔	✖	✖	✖
G_{49}(NH)-O_{80}(Hα)	✔	✔	✖	✖	✖	✖
G_{81}(NH)-O_{19}(Hα)	✔	✖	✖	✖	✖	✖

Table 2.2. Comparison of the observed ^{15}NH-^{15}NH and ^{15}NH-Hα NOE correlations in ABC-2 to those expected from the six possible registers. Check marks indicate that the expected cross peak is observed while the X marks indicate otherwise.

system. A second NOE between ^{15}NH- ^{15}NH could not be assigned due to spectral crowding. Since, ABC-1' and ABC-1 are similar except for the substitution of Hyp for Pro, these results provide very strong evidence for a four residue offset in ABC-1.**2.3 Conclusions**

The ABC-1 heterotrimer can putatively form 10 lateral salt- bridges while maintaining a canonical single-residue offset. Instead, our data suggest that the peptide chains in ABC-1 adopt a four-residue offset that sacrifices four hydrogen bonds at the termini in addition to significant van der Waals contact area to allow the formation of 10 axial salt-bridges. This demonstrates a clear energetic preference of axial over lateral salt-bridges. A similar preference for an axial salt-bridge has been demonstrated in systems where both contacts are feasible.[12] It has also been shown that Lys and Asp/Glu residues in lateral positions are conformationally frustrated due to favorable backbone polar contacts.[13] Although lateral interactions have been used in heterotrimeric design,[20,21] the results presented here add strong evidence of the dominance of axial interactions over lateral ones.

Characterization of supramolecular polymers is typically limited to relatively low resolution techniques such as AFM and TEM, leaving the exact mechanism of assembly vague and open to debate. This makes improvement and modification of these systems more difficult. Here we have created a model system of a sticky-ended collagen helix which, while having offset termini, does not continue assembling due to the relatively short, four- amino-acid stagger. This has allowed us to examine the structure by NMR and reveal the individual amino acid interactions responsible for the structural organization. Our study gives molecular-level evidence for axial salt-bridge-driven formation of a sticky-ended collagen triple helix for the first time. The design can be readily adapted to create collagen triple helices with longer offset ends that hybridize into higher order fibrillar assemblies. Failure to observe a lateral salt-bridge-driven canonical triple helix indicates that axial salt-bridges are better suited for controlling self-assembly of collagen triple helices.

The design of triple helices becomes an increasingly challenging topic as the number of possible alternative structures is considered. Moving from a homotrimer to a canonical ABC heterotrimer increases the number of reasonable structure from 1 to 27. An ABC system that makes non-canonical, sticky-ended structures feasible increases the number of competing structures to well over 100, even if only one triplet of overhang is considered. By increasing the length of the overhang, more competing states can be populated resulting in a rougher folding landscape, which in turn increases the time of folding. The effect of this is to roughen the folding landscape and therefore substantially increase the time of folding. From one perspective, the folding time may eventually make even well-designed systems impractical. From another perspective, in truly well-designed

systems, this additional challenge will be taken to heart, and negative design strategies will be used to engineer the folding landscape.

2.4 Experimental

ABC-1 and ABC-2 are identical with respect to the triple- helical core (see Scheme 2.1, shaded area), which facilitates comparison of their stability and molecular structure. All peptides were synthesized with an N-terminal Trp to determine accurate concentration and one ^{15}N-isotopically enriched Gly (^{15}N-Gly) per peptide for NMR analysis (Table 1). Individual peptides and their 1:1 binary and 1:1:1 ternary mixtures were examined via circular dichroism (CD) and NMR. The abbreviations and melting temperatures (T_m) of all peptide systems examined in this study are listed in Table 1.

2.4.1 Peptide Synthesis and Purification

Peptides were prepared with N-terminal Trp-Gly (WG) to assist in spectrophotometric determination of concentration. One ^{15}N- isotopically enriched Gly in each peptide chain is incorporated for NMR analysis. Sequence position of each residue is indicated by numeric subscripts starting from first amino acid of the leading chain. Therefore, N-terminal Pro of the leading, middle and lagging chains are labeled as P_3, P_{35}, and P_{67}, respectively.

All peptides were synthesized on an Advanced Chemtech Apex 396 solid phase peptide synthesizer using methods reported previously using 0.55 mM loading Fmoc-Rink Amide resin.[16] Briefly, 4:1 molar ratio of HATU and DIEA to amino acid coupling cycles were interleaved with deprotection cycles of 25% piperidine in DMF. For coupling of the isotopically enriched amino acid, the automated synthesis was stopped after the

deprotection cycle at the desired position and one equivalent of ^{15}N-isotopically enriched Fmoc-Gly-OH mixed with 1 equivalent of HATU and 1.5 equivalents of DIEA. All residues in the Xaa position were coupled twice to prevent deletions. The N-terminal of all peptides except A2 was left as a free amine. A2 was acetylated using a 0.7: 0.15: 5 volume ratio of acetic anhydride, DIEA, and DMF. A2 peptide is part of the ABC-2 helix designed to form a canonical collagen triple helix with a single residue offset between the chains. In this peptide arrangement, the electrostatically charged N-termini in case of free amine is placed in close vicinity inducing repulsion. It has been shown before that this electrostatic repulsion increases conformational disorder at the termini and decreases thermal stability. Additionally, the presence of free N-termini is also expected to slow kinetics of self-assembly. On the other hand, in our experience, tryptophan with N-terminal acetylation significantly reduces solubility of the peptides. Therefore, in order to strike a balance between solubility and conformationally disordered termini, we modified only one peptide in the ABC-2 system to have an N-terminal acetyl group. Since ABC-1 is designed to form a non-canonical triple helix with four residue offset, the electrostatically charged N-termini are placed far apart, preventing significant electrostatic repulsion. Therefore, acetylation of the N-terminal residues in ABC-1 was not required.

The peptides were cleaved from the resin using a mixture of trifluororoacetic acid, triisopropylsilane, water, and 1,2-ethanedithiol in a 37.6: 0.4: 1: 1 volume ratio. The cleaved peptide was triturated in cold diethyl ether and centrifuged to recover the peptide in a pellet form. The pellet was dissolved in mili-Q water and purified in reverse phase HPLC on a Varian PrepStar220 with a preparative C-18 column. HPLC was carried out

in water with a 1% v/v gradient of acetonitrile in presence of 0.05% trifluoroacetic acid. The fractions collected with HPLC were then analyzed using electrospray ionization time-of-flight mass spectroscopy (ESI-TOFMS) on a Bruker ESI-microTOF. The HPLC chromatogram and mass spectra are available in Figs. 2.8 and 2.9.

The HPLC purified peptides were rotary-evaporated to remove acetonitrile and lyophilized to recover a white powder that was used in preparation of CD and NMR samples.

2.4.2 Preparation of Samples for CD and NMR

All CD and NMR experiments were performed at 1 mM and 3 mM total peptide concentration, respectively, in presence of 10 mM TRIS buffer at pH 7.2 *in absence* of salts. In case of ABC-1, peptides A1, B and C1, their 1:1 binary mixtures A1:B, C1:B and A1:C1 and 1:1:1 ternary mixture A1:B:C1 were examined via CD and NMR. Similarly, all peptides and their combinations possible for ABC-1 and ABC-2 systems were also examined. Stock solution of each peptide was prepared by dissolving ~10 mg of lyophilized powder in 1ml of MiliQ- water. The pH of stock solutions was adjusted with 1M NaOH to neutral and their concentration determined spectrophotometrically on a Nanodrop 1000 using tryptophan molar extinction coefficient of 5502 M^{-1} cm^{-1} monitored at 280 nm. 3 mM solution of individual peptides, their 1:1 binary or 1:1:1 ternary mixtures were prepared by diluting the stock solutions prepared above with appropriate volumes of MiliQ-water and 100 mM TRIS buffer at pH 7.2 in absence of any salts. The prepared samples were heated to 85 °C for 5 minutes, and then left to incubate for 1 week at room temperature. In order to prepare samples for CD experiments, incubated 3mM stock solutions were diluted with 10 mM TRIS buffer at pH

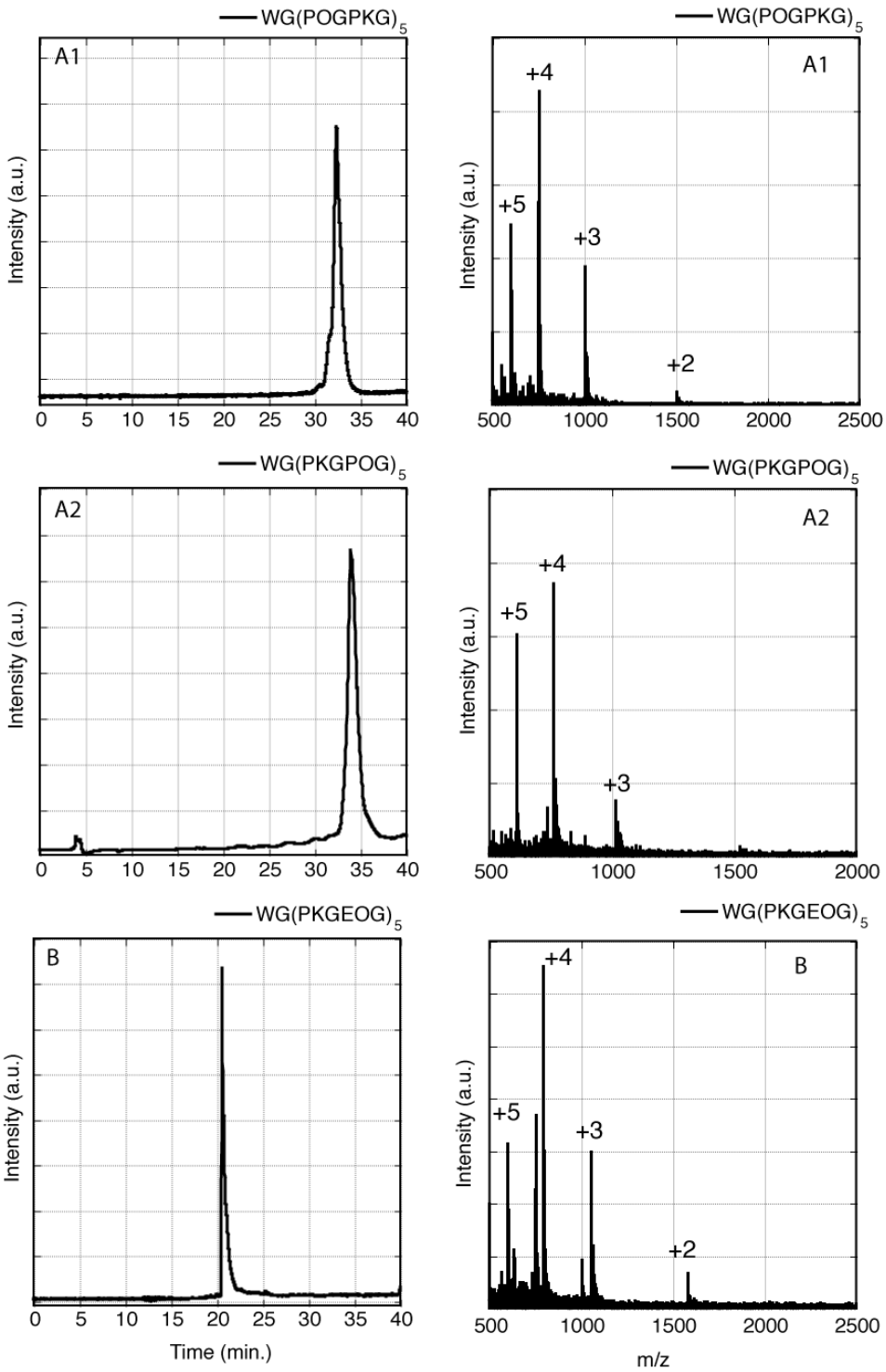


Figure 2.7. HPLC chromatogram (left) and ESI-MS spectra of peptides A1) $WG(POGPKG)_5$ monoisotopic mass$[M+2H]^{2+}$, expected: 1504.8 m/z, observed: 1504.8 m/z., A2) $WG(PKGPOG)_5$ monoisotopic mass$[M+3H]^{3+}$, expected: 1017.5 m/z, observed: 1017.6 m/z and B) WG(PKGEOG)5 monoisotopic mass $[M+2H]^{2+}$, expected: 1584.8 m/z, observed: 1584.8 m/z.

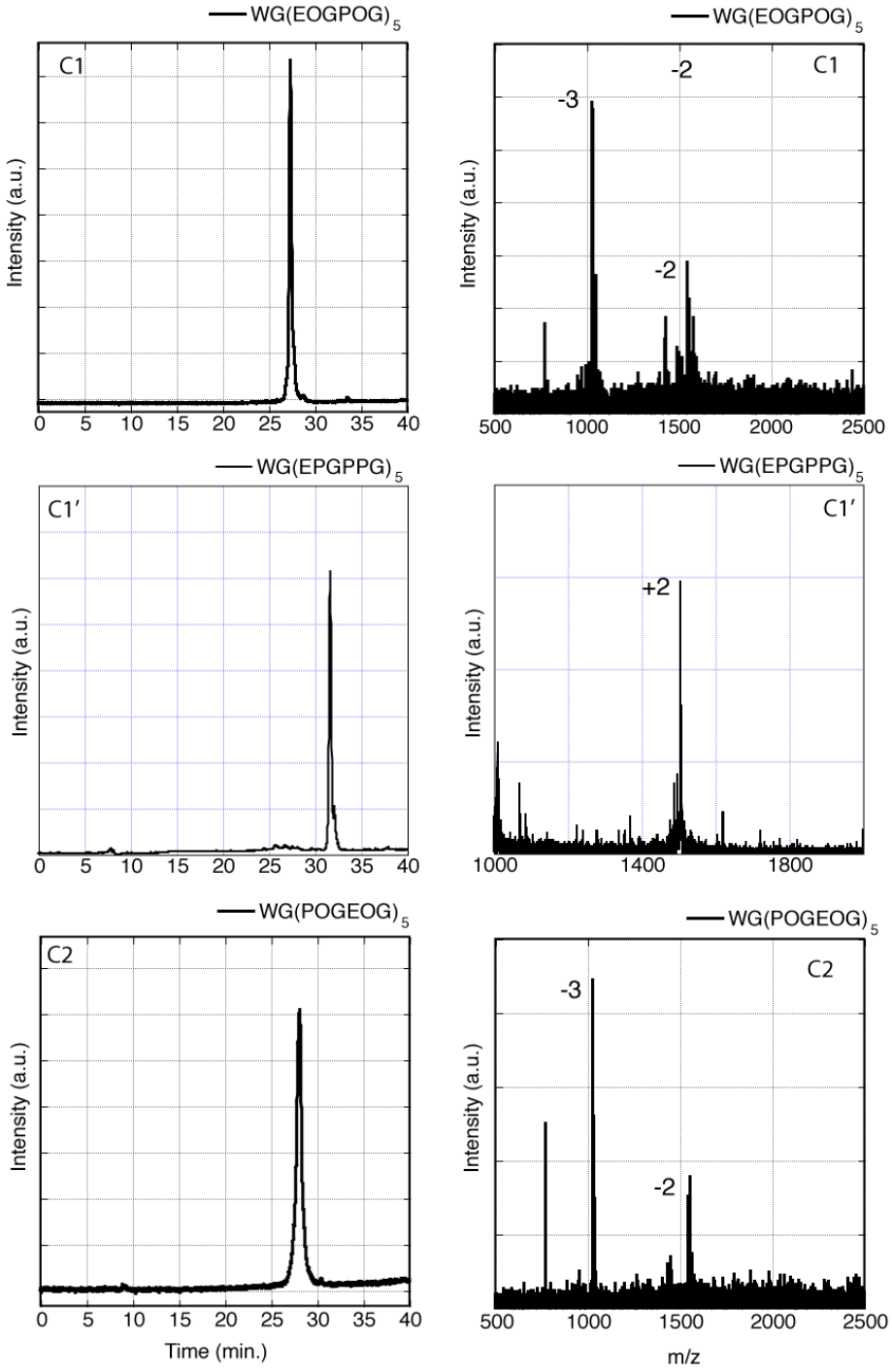


Figure 2.9. HPLC chromatogram (left panel) and ESI-MS (right panel) of peptides. C1) $WG(EOGPOG)_5$ monoisotopic mass $[M\text{-}2H^-]^{2-}$, expected: 1545.1 m/z, observed: 1545.3 m/z, C1') $WG(EPGPPG)_5$ monoisotopic mass $[M+2K]^{2+}$, expected: 1505.1 m/z, observed: 1505.4 m/z and C2) $WG(POGEOG)_5$ monoisotopic mass $[M\text{-}2H]^{2-}$, expected: 1545.1 m/z, observed: 1545.3 m/z.

7.2 to obtain 1 mM peptide solutions. This sample was then left at 5 °C for 8 hours or overnight with the exception of the peptide combinations associated with ABC-2 which was incubated at 5 °C for 1 week. Longer incubation times were needed for these peptides to obtain high quality thermal melting transitions. After the full incubation, 45 μL of solution was loaded into a 0.1 mm path length Hellma QS quartz cuvette and transferred to spectropolarimeter sample chamber for measurement.

2.4.3 Circular Dichroism Analysis

All CD experiments were performed on a Jasco J-810 spectropolarimeter with a Peltier temperature-controlled stage. For each sample, two different experiments were performed. First, a wavelength scan between 250 and 190 nm at a scan rate of 50 nm/min and a data pitch of 0.1 nm was performed to determine the secondary structure characteristic of species present in solution. In all cases, a maximum was observed between 223-226 nm. In the second experiment, a wavelength-temperature scan was performed in which ellipticity (θ) was monitored as a function of temperature (T) at the maximum observed in wavelength scan experiments. Temperature was ramped from 5 to 85 °C at a scan rate of 10 °C/hour and a data acquisition pitch of 0.5 °C The molar residual ellipticity (MRE) of each samples was calculated according to the following equation:

$$[\theta] = \frac{\theta}{c \times l \times nr \times 10}$$

where **θ** is the observed ellipticity in millidegrees observed in thermal unfolding experiments, c is the total concentration of peptide in moles l^{-1}, l is the path length in cm, and nr is the number of residues in the individual peptide. First derivative (dθ/dT) of the melting curves was calculated using a Savitzky-Golay smoothing algorithm. The

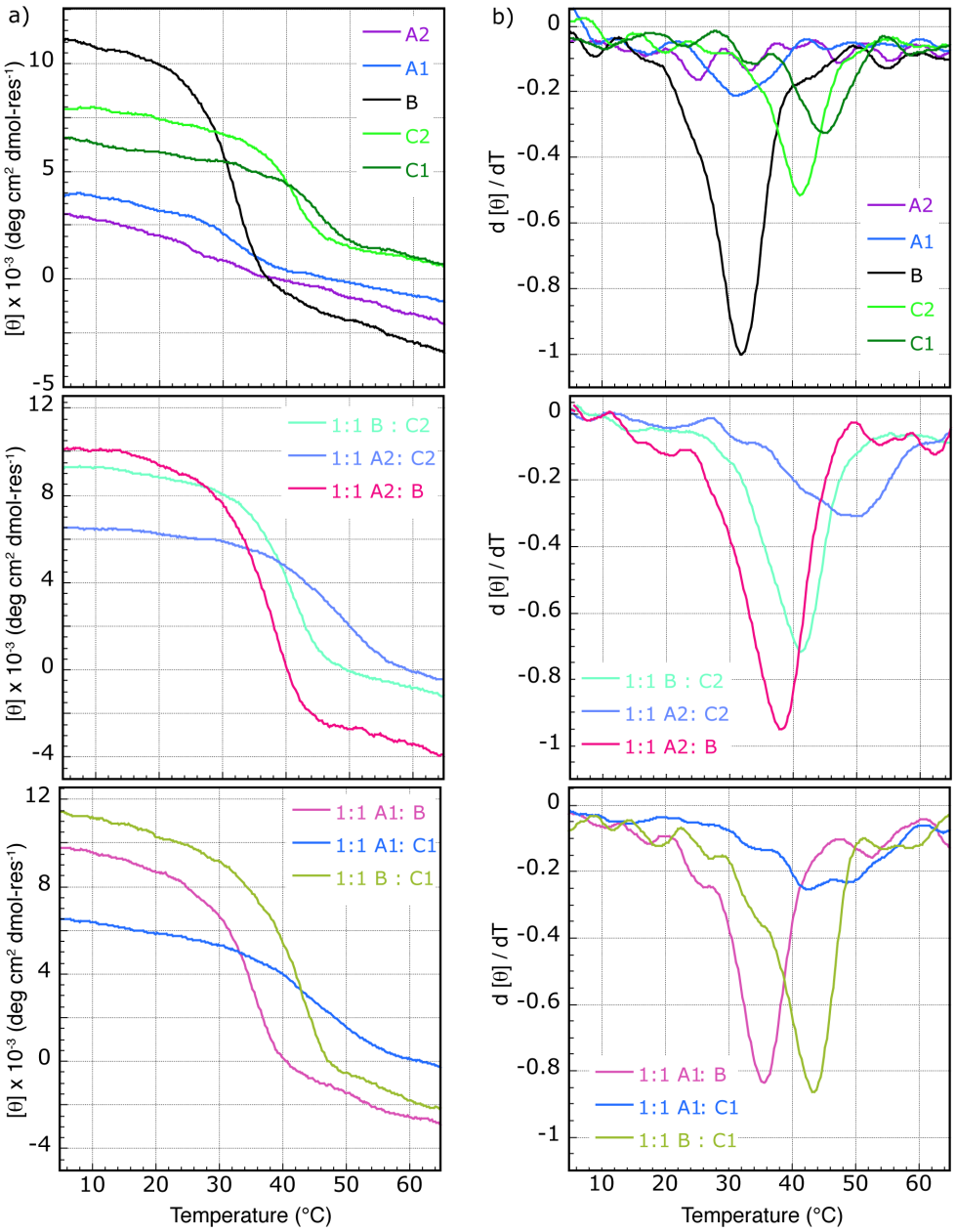


Figure 2.9. Thermal melting profile (A) and their first derivative curves (B) of individual and binary mixtures of all peptides used in this study.
ABC-2 system are provided in figure 2.10.

minimum observed in the dθ/dT plot is reported through the text as melting temperature, T_m. The melting curves and first derivative plots of peptide combinations in ABC-1and

2.4.4 Acquisition and processing parameters in NMR spectroscopy

The ^{1}H,^{15}N-HSQC, ^{1}H,^{1}H-TOCSY, ^{1}H,^{1}H- NOESY and a two dimensional ^{1}H-^{1}H

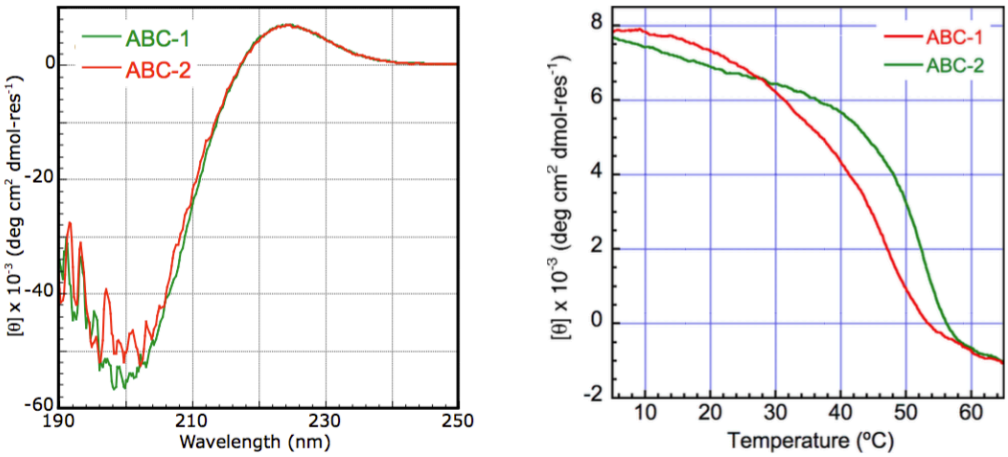


Figure 2.10. CD wavelength scan (left panel) and thermal melting profile (right panel) of ABC-1 (green) and ABC-2 (red). The wavelength scans of both ABC heterotrimers show similar CD spectra with a positive ~ 225 nm maxima and a negative ~200 nm minimum characteristic of the left-handed poly-proline type II peptide conformation in the triple helix. A cooperative thermal unfolding is observed for both heterotrimers in the right panel.

version of the three dimensional $^{1}H,^{15}N$-NOESY-HSQC spectra of all peptide combinations in ABC-1 and ABC-2 system were acquired on Varian Inova Spectrometer operating at a ^{1}H frequency of 500 or 600 MHz at 26 °C or 40 °C, unless otherwise noted. Due to significant overlap in the amide region, a three dimensional $^{1}H,^{15}N$-NOESY-HSQC spectra of the ABC-1 heterotrimer was acquired on a Bruker cold probe operating at ^{1}H frequency of 750 MHz. In all cases, the ^{1}H and ^{15}N carrier frequencies were set to H_2O resonance and 108 ppm, respectively.

A fast version of the $^{1}H,^{15}N$-HSQC experiment was used to acquire two dimensional amide spectra with 2398 (F2) x 256 (F1) complex points in 8 scans. The ^{1}H and ^{15}N dimension acquisition windows were set to 12 and 15 ppm, respectively. Homonuclear $^{1}H,^{1}H$-TOCSY spectra was acquired with 2048 (F2) x 512 (F1) complex points in 16 scans with a mixing time of 50 or 75 ms and a spin lock of 8000 Hz. The $^{1}H,^{1}H$-NOESY spectra was acquired with 2048 (F2) x 1024 (F1) complex points in 16 scans with a mixing time of 125 ms. A 12 ppm and 10 ppm spectral window in the direct

and indirect dimension, respectively was used for both TOCSY and NOESY experiments. A two dimensional $^{1}H,^{1}H$ version of a 3D $^{1}H,^{15}N$-NOESY-HSQC experiment was acquired with 2048 x 256 complex points in 256 scans with a mixing time of 100 and 125 ms and a spectral window of 12 (F2) x 10 (F1) ppm. The ^{15}N dimension evolution time was held constant during the experiment to acquire only the homonuclear plane. The 3D $^{1}H,^{15}N$-NOESY-HSQC spectrum of ABC-2 was acquired with 1792 (F2) x 24 (F3) x 256 (F1) complex points in 128 scans at 125 ms mixing times. The ^{1}H and ^{15}N carrier frequencies were set to water resonance and 108 ppm, respectively. A spectral window of 10 ppm was used for both proton dimensions and a spectra window of 8 ppm was used for the nitrogen dimension.

All acquired spectra were processed using NMRPipe[22] and analyzed in CcpNmr[23]. Appropriate solvent suppression filter followed by cosine-squared bell apodization was applied to both dimensions and the data was zero filled to the next power of two and Fourier transformed. Forward and backward linear predictions were used to extend the time domain data of the homonuclear experiments, wherever necessary. Chain specific sequential assignment assignments were carried out as reported previously.[16]

2.5 References

(1) Rothemund, P. W. K. *Nature* **2006**, *440*, 297-302.

(2) Shih, W. M.; Quispe, J. D.; Joyce, G. F. *Nature* **2004**, *427*, 618-627.

(3) Goodman, R. P.; Berry, R. M.; Turberfield, A. *J. Chem. Commun.* **2004**, 1372-1373.

(4) He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. D. *Nature* **2008**, *452*, 198-201.

(5) Gradisar, H.; Bozic, S.; Doles, T.; Vengust, D.; Hafner-Bratkovic, I.; Mertelj, A.; Webb, B.; Sali, A.; Klavzar, S.; Jerala, R. *Nat. Chem. Biol.* **2013**, *9*, 362-366.

(6) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell,

J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science* **2013**, *340*, 595-599.

(7) Pandya, M. J.; Spooner, G. M.; Sunde, M.; Thorpe, J. R.; Rodger, A.; Woolfson, D. N. *Biochemistry* **2000**, *39*, 8728-8734.

(8) Gribbon, C.; Channon, K. J.; Zhang, W. J.; Banwell, E. F.; Bromley, E. H. C.; Chaudhuri, J. B.; Oreffo, R. O. C.; Woolfson, D. N. *Biochemistry* **2008**, *47*, 10365-10371.

(9) Kotch, F. W.; Raines, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 3028-3033.

(10) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. *Science* **1994**, *266*, 75-81.

(11) Okuyama, K.; Miyama, K.; Mizuno, K.; Bachinger, H. P. *Biopolymers* **2012**, *97*, 607-616.

(12) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(13) Fallas, J. A.; Dong, J. H.; Tao, Y. Z. J.; Hartgerink, J. D. *J. Biol. Chem.* **2012**, *287*, 8039-8047.

(14) O'Leary, L. E. R.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, *J. D. Nat. Chem.* **2011**, *3*, 821-828.

(15) O'Leary, L. E.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2011**, *133*, 5432-5443.

(16) Jalan, A. A.; Hartgerink, J. D. *Biomacromolecules* **2013**, *14*, 179-185.

(17) Fallas, J. A.; Hartgerink, J. D. *Nat. Commun.* **2012**, *3*, 1-8.

(18) Jalan, A. A.; Demeler, B.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2013**, *135*, 6014.-6017

(19) Fallas, J. A.; Lee, M. A.; Jalan, A. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2012**, *134*, 1430-1433.

(20) Xu, F.; Zahid, S.; Silva, T.; Nanda, V. *J. Am. Chem. Soc.* **2011**, *133*, 15260-15263.

(21) Xu, F.; Zhang, L.; Koder, R. L.; Nanda, V. *Biochemistry* **2010**, *49*, 2307-2316.

(22) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.

(23) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, P.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 687-696.

Chapter 3: AAB Heterotrimers to Investigate Mutations of the B Chain*

3.1 Introduction

Collagen is a triple helical structure consisting of three strands with an (Xaa-Yaa-Gly)$_n$ repeat, and is naturally found as both a homotrimeric composition (where all three chains are identical) and heterotrimeric compositions (where one or two peptide chains are sequentially unique). Each chain is offset by one amino acid, allowing a glycine residue to be present in every cross section of the triple helix. Glycines fold into the center of the helical axis and form a stabilizing hydrogen bond with a neighboring chain, creating the unique triple helical fold.[1,2] The offset between peptide strands introduces a register, or relative order (eg: ABC vs. CBA) of the strands in the folded helix. Though not evident in homotrimers, register becomes apparent in heterotrimers where each arrangement of strands results in a distinct three-dimensional surface recognized by receptors and other macromolecules.

The methodology now exists to rationally create specific single composition and register triple helices[3-5] utilitzing axial charged pairs, which create an energetic benefit for chains in the correct geometry.[6] Many collagens, including the abundant collagen type I, are known to be heterotrimeric.[7] With the advancement of heterotrimeric design, it is now possible create more realistic models of disease phenotypes. Particularly of interest

---

* The work in this chapter was published as Acevedo-Jake, A.M.; Clements, K.A.; Hartgerink, J.D. "Synthetic, Register-Specific, AAB Heterotrimers to Investigate Single Point Glycine Mutations in Osteogenesis Imperfecta" *Biomacromolecules*. **2016** *17*, 914−921

AAJ and KAC synthesized and purified the peptides. AAJ performed NMR experiments. KAC performed CD and modeling experiments. AAJ, KAC, and JDH analyzed the data and wrote the manuscript.

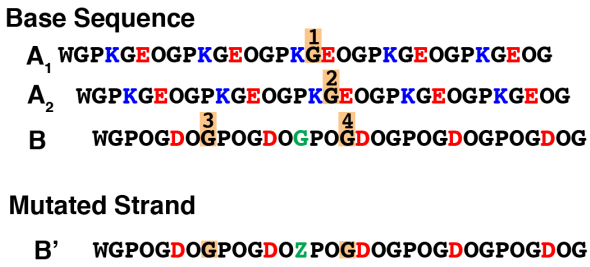


Figure 3.1. Sequence of A and B peptides used. Green represents the mutation site, blue amino acids are positively charged and red amino acids are negatively charged. Boxed amino acids are ^{15}N-isotopically labeled glycines and are numbered for easy correlation to NMR data.

is the disease *Osteogenesis Imperfecta* (OI), which typically results from single point amino acid mutations in the collagen sequence. In OI a mutation substitutes one of the obligate glycines for another amino acid and disrupts the folding pattern of the triple helix.[8-10] This has significant consequences on the helical structure because these amino acids are larger than glycine and interrupt the canonical hydrogen bonding pattern.

The 1CAG crystal structure was the first triple helix with an OI type mutation to be crystallized.[11] The 1CAG structure consists of a homotrimer of $(POG)_5POA(POG)_4$ peptides and shows significant disruption to the hydrogen bonding pattern. For example, four of the expected hydrogen bonds around the mutation site are replaced with water-mediated hydrogen bonds. This disruption widens the diameter of the triple helix to 4.5 Å, from a normal maximum near 3 Å. Other studies have performed NMR analysis on this and similar homotrimeric[12-15] systems as well as studies involving type IV collagen.[16,17] For the 1CAG peptide, the crystal structure showed more hydrogen bonding disruption than that observed in the solution state by NMR, where the alanine in the leading chain participates in a hydrogen bond. Nevertheless, these structures clearly show the impact of glycine mutations on the triple helix.

Our lab previously designed a heterotrimeric sequence with an AAB guest region of known OI mutations.[18] Multiple combinations of mutations were tested in both the A and B chains in order to obtain mutant helices with one, two or three glycine mutations. That study found that each mutation introduced had a decreasing impact on the stability of the helix. The first mutation had the largest effect on the melting point of the helix, with each mutation thereafter having a smaller effect. While that study gave clear information about the melting point of the mutated triple helices, no information about the structure could be obtained because the system was not register or composition specific.

Now that the methodologies exist to create robust single register, single composition heterotrimeric triple helices, here we have designed a system with composition and register control containing a single OI-like glycine substitution. The overall design is based on a previously published controlled register and composition AAB heterotrimer,[3] shown in Figure 3.1. This triple helix has a high melting point of 51 °C, which allows for some of the destabilized mutated helices to fold at or above room temperature. Amino acids were selected to cover a wide range of naturally presented OI mutations including charged, branched, hydrophobic, and polar amino acids. The structure and composition of these systems was characterized by CD and NMR, and from these analyses distance constraint based models were generated to better visualize molecular level structural disruptions.

3.2 Characterization of Mutant Helices

3.2.1 Base System

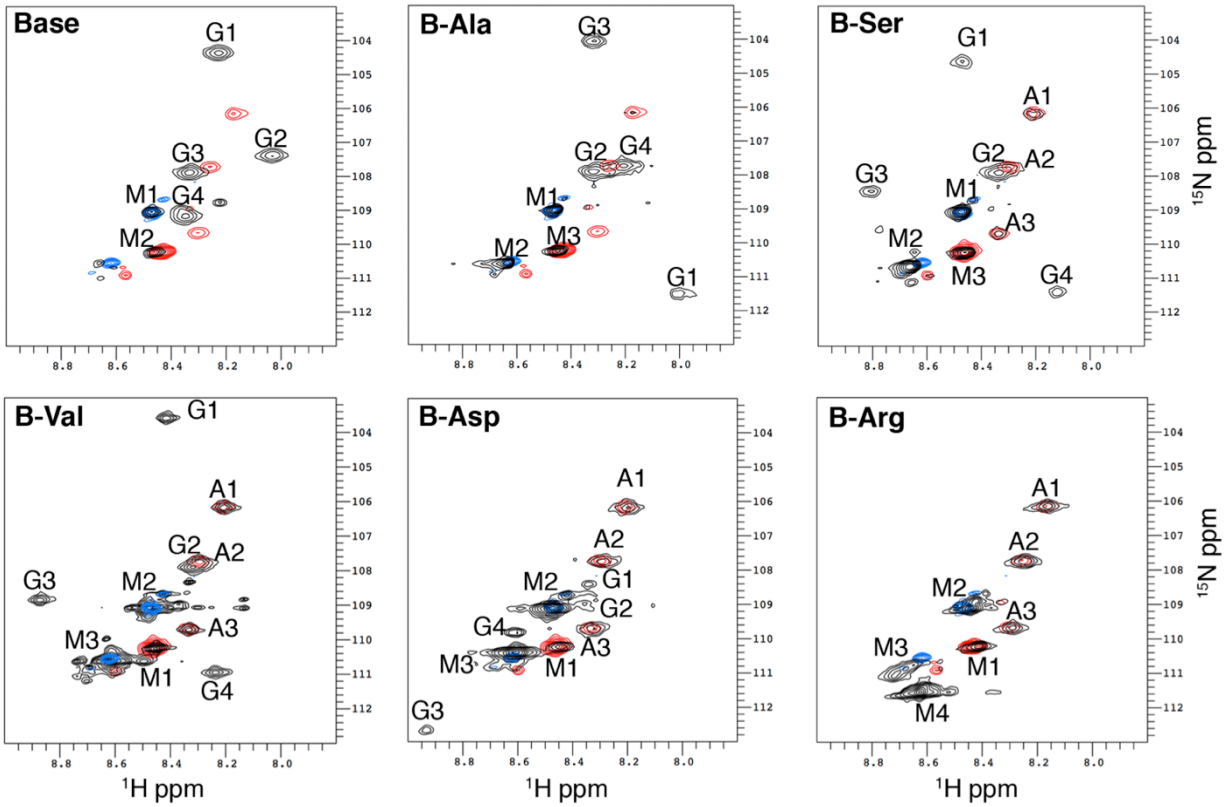


Figure 3.2. ^{1}H,^{15}N-HSQC spectra of base and mutant systems. Blue traces show the B' mutant peptide alone, red traces show the A peptide alone. Black shows the 2:1 AAB' mixture of the peptides. Peaks labeled M represent monomer, those labeled A represent A homotrimer while peaks labeled G represent the AAB' mutant labels as labeled in Fig. 1

To reflect the composition of type I collagen, a previously designed[3] AAB single composition and single register system was employed as the template to model single site *Osteogenesis Imperfecta* mutations. Isotopically^{15}N-enriched glycine was incorporated into the B and A chains (Figure 3.1) to verify composition of the assembled helix by ^{1}H,^{15}N-HSQC spectroscopy.

As previously observed,[3] because of a high number of unsatisfied charged residues the B homotrimer does not fold and we record only monomeric peptide by both CD and NMR. The A peptide, however, can fold into a homotrimer (three peaks of equivalent volume by ^{1}H,^{15}N-HSQC) with a melting temperature of 32 °C. The 2:1 mixture of the A and B peptides results exclusively in a single register and single composition AAB

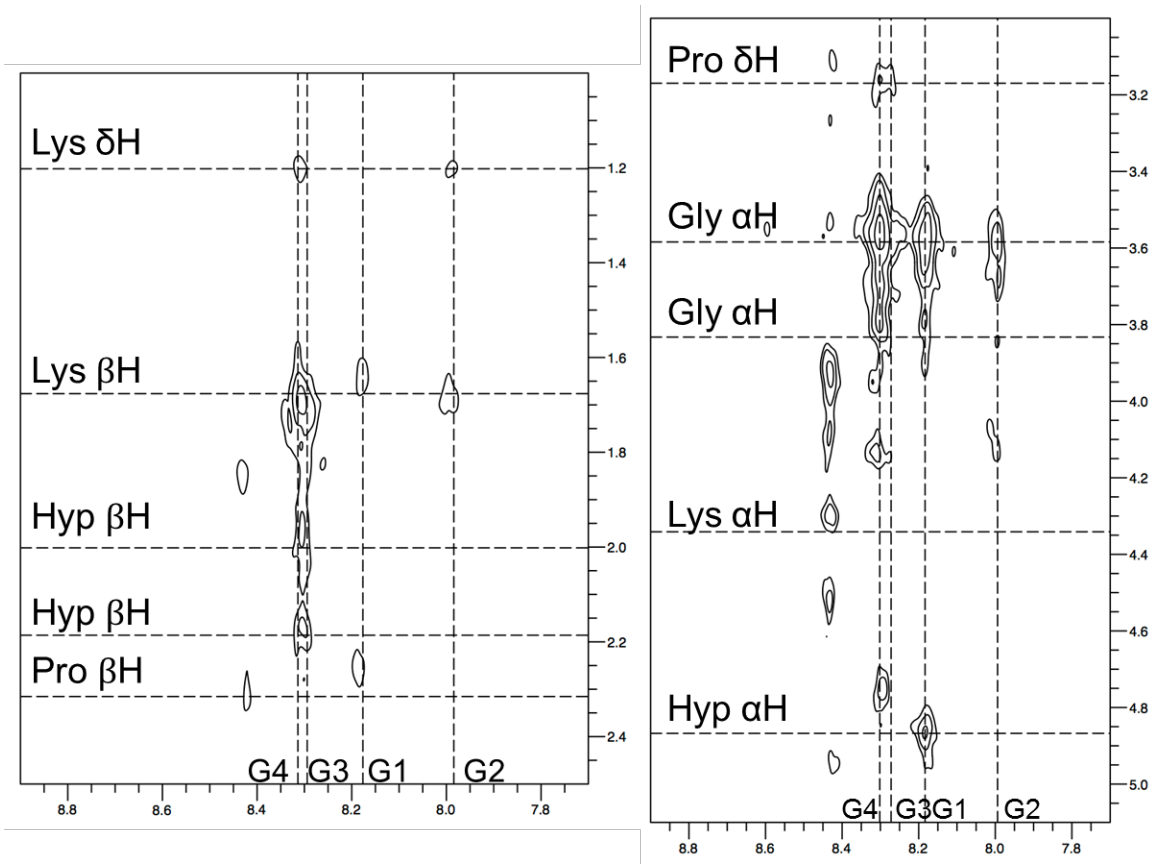


Figure 3.3. NOESY HSQC of the AAB base peptide.

heterotrimer. This system displays a characteristic helical signal by CD spectroscopy with a maximum around 225 nm and a minimum at 190 nm, and has a high melting temperature of 51 °C. By NMR, four peaks (G1, G2, G3 and G4) of equal volume in the $^{1}H,^{15}N$-HSQC spectra confirm AAB composition (Figure 3.2). Comparison of the folded AAB peak volume to the monomeric peptides (M1 and M2) reveals the equilibrium population of base trimer is high in comparison to the mutant systems (Table 3.1). Additionally, measurement of T1 and T2 relaxation values for the isotopic labels was performed. In agreement with previous reports, the monomeric single peptide strands are characterized by a low R2/R1 ratio (around 2.00 or below), while the folded helical peaks show significantly higher R2/R1 ratios (Table 3.3). Finally these ^{15}N-glycine labels were

Alpha homotrimer peak	*T1*	*T2*
A1	*0.459±0.031*	*0.032±0.008*
A2	*0.522±0.042*	*0.050±0.009*
A3	*0.399±0.041*	*0.060±0.011*

Table 3.1. Alpha homotrimer relaxation values.

Peak	*Base*	*B-Ala*	*B-Ser*
G1	*14.85*	*20.76*	*15.72*
G2	*9.55*	*10.67*	*13.53*
G3	*21.32*	*13.79*	*18.63*
G4	*8.00*	*10.02*	*17.84*

Table 3.2. Peak volumes from ^{1}H, ^{15}N-HSQC

used to probe the surrounding helical topology by ^{1}H,^{15}N-NOESY-HSQC. The NOEs determined for each label (Appendix 1) were used to generate a model of the base heterotrimer. These data characterize the base AAB single register single composition system, and are used as a reference of comparison to the mutant systems.

3.2.2 Mutant Systems

For simplicity, AAB heterotrimeric helices with a glycine substitution in the B chain are named after the amino acid which replaces glycine, for example B-Ala or B-Ser. The A chain is identical in all cases. With the single mutation in the helix, our series shows increasing disorder associated with the character of amino acid (Ala > Ser > Val > Asp > Arg) which, with the exception of arginine, closely resembles the trend shown the clinical cases[8,24] and previous reports.[25] The mutations are ordered using a combination of the melting points from CD and the composition information from NMR experiments. Unlike previous reports however, we are able to report structural as well as compositional disruptions to the initial base system. Additionally, in the case of B-Ala and B-Ser, we prepared NOE guided models to visualize these localized disruptions.

3.2.2.1 Alanine

Substitution	Name	AAB' T_m	Folding	R2/R1 of AAB'	Monomer Rel Vol.	AAB' Rel. Vol.	AAA Rel. Vol.
Glycine	BASE	51 °C	Single composition, single register AAB	14.85, 9.55, 21.32, 8.00	0.33	0.67	-
Alanine	B-Ala	30.5 °C	Single composition, single register AAB'	20.76, 10.67, 13.79, 10.02	0.44	0.56	-
Serine	B-Ser	29.5 °C	AAB' folded as well as AAA homotrimer	15.72, 13.53, 18.63, 17.84	0.61	0.37	0.02
Valine	B-Val	-	Multiple compositions	10.28, 9.62, 7.84, 4.71	0.63	0.21	0.16
Aspartate	B-Asp	-	Partially Folded	7.10, 5.24, 2.11, 8.35	0.75	0.07	0.18
Arginine	B-Arg	-	Only homotrimers	-	0.74	-	0.26

Table 3.3. Summary of experimental properties of the sequences. Melting points are from Circular Dichroism. Relaxation values are from ^{15}N relaxation experiments on the heterotrimeric solutions in numerical order of the label. Peak volumes are from the ^{1}H,^{15}N-HSQC spectra of the mixed heterotrimeric solutions. A more detailed description of the assignment process, additional relaxation values, and TOCSY are given in Appendix 2.

Of all the B chain mutations investigated, only the B-Ala system displays perfect composition and register control, on the basis of the single transition in CD combined with the single triple helical population in NMR studies. The CD melting spectra of this sample shows a single strong cooperative transition again suggesting this AAB' mutant heterotrimer folded (Figure 3.4) because there are no other species present by NMR.

Specifically, the ^{1}H,^{15}N-HSQC of the B-Ala reveals an AAB' heterotrimer similar to the base, with four peaks of roughly equal ratio (G1, G2, G3 and G4) and high R2/R1 ratios (Figure 3.2, Table 3.3). Comparison of equilibrium peak volumes of the folded versus unfolded peptides reveals the B-Ala AAB' population is only slightly lower than the base system (base is 0.33:0.67 monomer to trimer versus 0.44:0.56 in B-Ala), suggesting the energetic penalty of this mutation is quite mild. High relaxation values for the B-Ala mutant reveals that the labeled ^{15}N-glycines surrounding the mutation site are folded and well-ordered (Table 3.3). Because these values are more characteristic of organized helical rather than unfolded monomeric peptide, the structural disruptions to

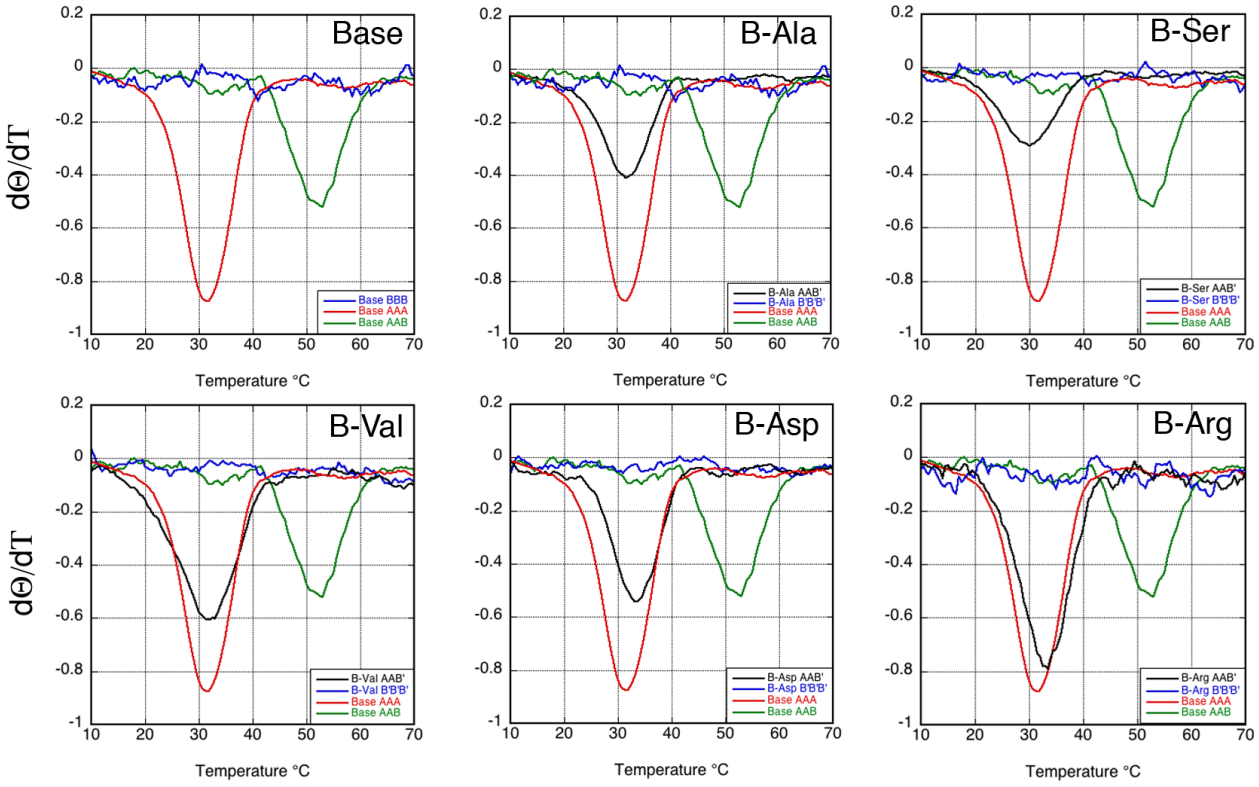


Figure 3.4. First derivative of Circular Dichroism melting experiments demonstrating the relative thermal stabilities of the compared systems. For easy comparison, the component peptides AAA and BBB and their combined AAB triple helix are shown in all graphs in red, blue and green respectively. Modified peptides with substitutions away from Glycine are shown in black.

the helical topology must be quite localized. This differs from some previous reports employing homotrimers, which report that the triple helix is not able to re-nucleate past the mutation site and the terminal regions remain unfolded.[26] Finally, because each of the labeled residues in the AAB' mutant lie at unique ^{1}H ppm values, it was possible to determine NOE correlations to surrounding protons for each isotopic label. These correlations (Appendix 1) were used to generate a model of the topology surrounding each of these amino acids to better visualize this comparison (described below).

3.2.2.2 Serine

The incorporation of a serine mutation into our sequence still allows triple helical folding, though we observe additional compositional disruptions. For the B-Ser system the AAA homotrimer and the AAB' heterotrimer are both observed by NMR to fold, but the AAA population is low when compared to the equilibrium population of mutant heterotrimer and monomer (0.02 : 0.37 : 0.61; homotrimer : heterotrimer : monomer). Overall, this results in less A chain available for heterotrimer folding, and reduces the total amount of folded AAB' helix present. In addition, the equilibrium ratio between folded and unfolded peptides is relatively lower than for the AAB base (Table 3.3). Relaxation values, as for the B-Ala mutation, confirm both that the triple helix re-nucleates on both sides of the mutation and that these regions are well-ordered

The remaining mutants, Val, Asp, and Arg show more significant compositional disruptions with larger relative amounts of homotrimer. Additionally, dynamic values of the mutant heterotrimers suggest more significant structural disruptions and decreased helical organization. In these cases, the system accommodated the mutation by minimizing the amount of folded mutant AAB' helix, shifting the equilibrium toward the AAA homotrimer. For these systems, signal overlap prevented unambiguous assignment of neighboring NOEs and helical register.

3.2.2.3 Valine

The valine mutation results in folding of the intended mutant heterotrimer in addition to a significant population of the A homotrimer. The CD melting profile shows a broader transition representing this mixed population (Figure 3.4). By NMR, the four $^{1}H,^{15}N$-HSQC peaks for the mutant have equal volumes and represent the AAB' composition. Similar to the B-Ala and B-Ser systems, the relaxation values confirm that

Triple Helix	Lost hydrogen bonds	Lost charge pairs	Compensating Interactions	Peak diameter of helix
B-Ala	5	3	8	4.7 Å
B-Ser	2	4	6	3.7 Å
1CAG [11]	4	N/A	6	4.2 Å
T1-898(G109S) [12]	4	N/A	N/A	~5 Å

Table 3.4. Comparison of lost and gained interactions. Lost hydrogen bonds are backbone hydrogen bonds that would be expected to take place in the Base structure, but no longer are formed in the same way. Compensating interactions are any interactions that are not present in the Base structure.

the helix refolds around the mutation site. Three of these peaks, G1, G2, and G3, have high R2/R1 ratios comparable to the base and must be well-ordered. The fourth peak, G4, has a lower R2/R1 ratio; though we still classify it as helical, this site is somewhat less organized than the others. This suggests that in the case of valine mutants the structural disruptions are less localized and propagate further along the helical backbone. The triple helix - monomer equilibrium is comparable to B-Ser system (Table 3.3), and slightly less favorable than for the base population. Unlike B-Ser however, the B-Val AAB' helix and the base AAA homotrimer populations are present in equal amounts (Table 3.3). Unfortunately for the B-Val system, signal overlap in the ^{1}H dimension prevented unambiguous NOE assignments to be made and therefore a model was not generated.

3.2.2.4 Aspartate

With the B-Asp mutation, the total equilibrium population of folded helices (homotrimers and heterotrimers) is low in comparison to the B-Ala, B-Ser, and B-Val systems (Table 3.3). Because the population of heterotrimers is so low, by CD only the strong AAA homotrimer transition is observed. However, by ^{1}H,^{15}N-HSQC it is possible to clearly distinguish both the B-Asp heterotrimer and the AAA base homotrimer. For

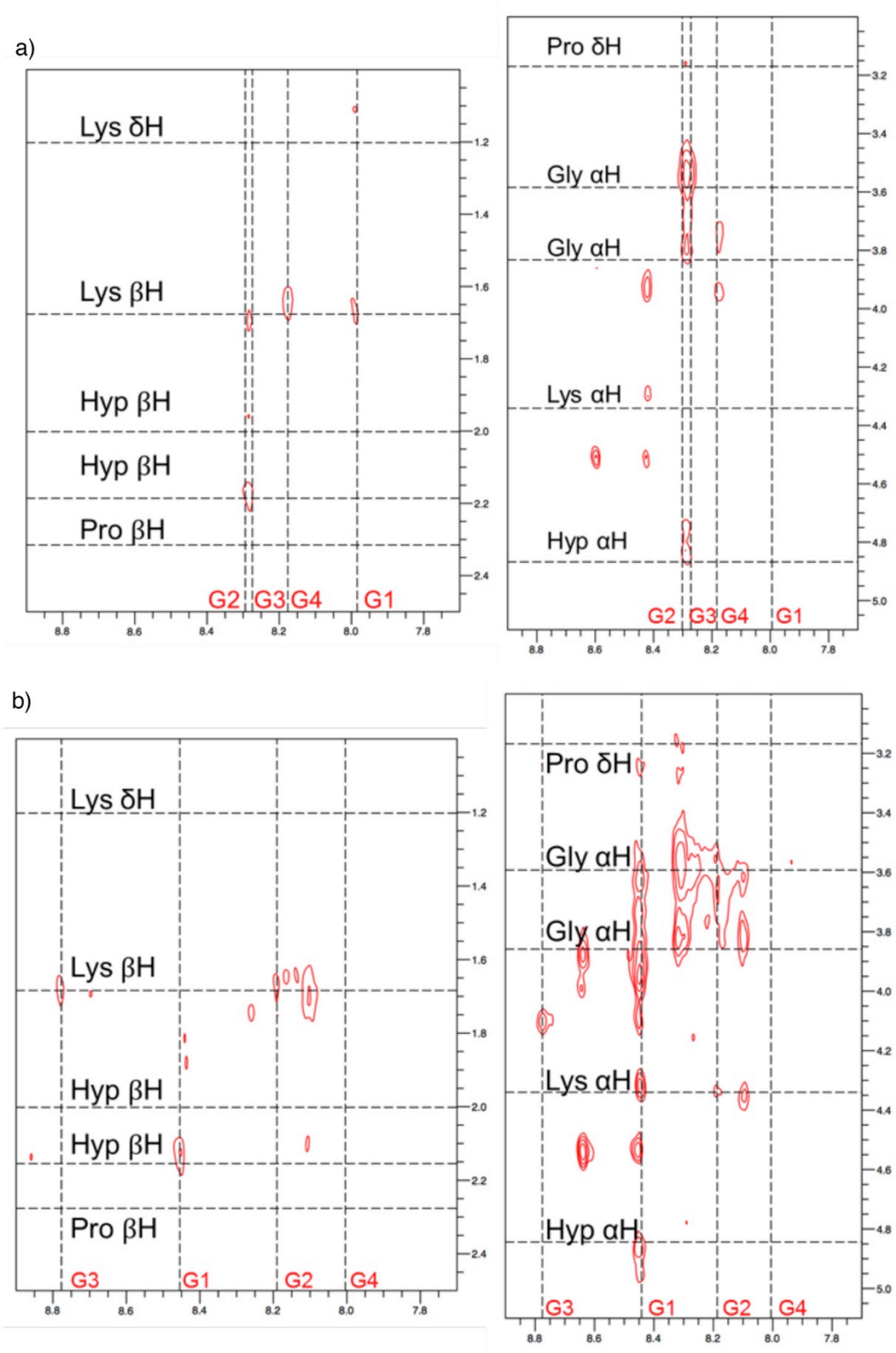
this B-Asp system however, the homotrimer peak volume is four times greater than B-Asp helix peak volume, corroborating the phenotypic severity reported from previous studies associated with aspartate mutations.[25] The four AAB' $^{1}H,^{15}N$-HSQC peaks, G1, G2, G3 and G4, show equal volumes; for three of these peaks, G1, G2, and G4, the R2/R1 values classify them as ordered triple helix. The fourth heterotrimer peak, G3, has a lower R2/R1 value and is not well-ordered. This relaxation data reveals that, like the valine mutant, the helical backbone surrounding the aspartate mutation is less well-organized than the B-Ala and B-Ser systems.

3.2.2.5 Arginine

Previous studies of OI mutations have shown that charged and branched amino acid mutations result in the harshest patient phenotypes and largest disruptions to the helical structure.[8,9,25] Consistent with these reports, characterization of the B-Arg system shows the most perturbed composition of all the prepared mutants. The AAA base homotrimer, in addition to the monomers for both the A and B' chains are apparent in NMR (Figure 3.2). For this system there were no definitively assignable populations of AAB' heterotrimer. The presence of the new and large monomeric peak, M4, suggests a partially folded intermediate. The considerable steric and conformational consequences of incorporating arginine into the B chain apparently prevents heterotrimer formation, and instead the equilibrium favors the monomer state and AAA homotrimer.

3.3 Structural Disruptions around Mutation Site

Both the alanine and serine mutant systems displayed well-folded AAB' heterotrimeric populations, where the register and composition was clear and unambiguous. Because each of the labeled residues for both mutants lay at unique ^{1}H



.**Figure 3.5**. NOESY-HSQC of B-Ala (a) and B-Ser (b).

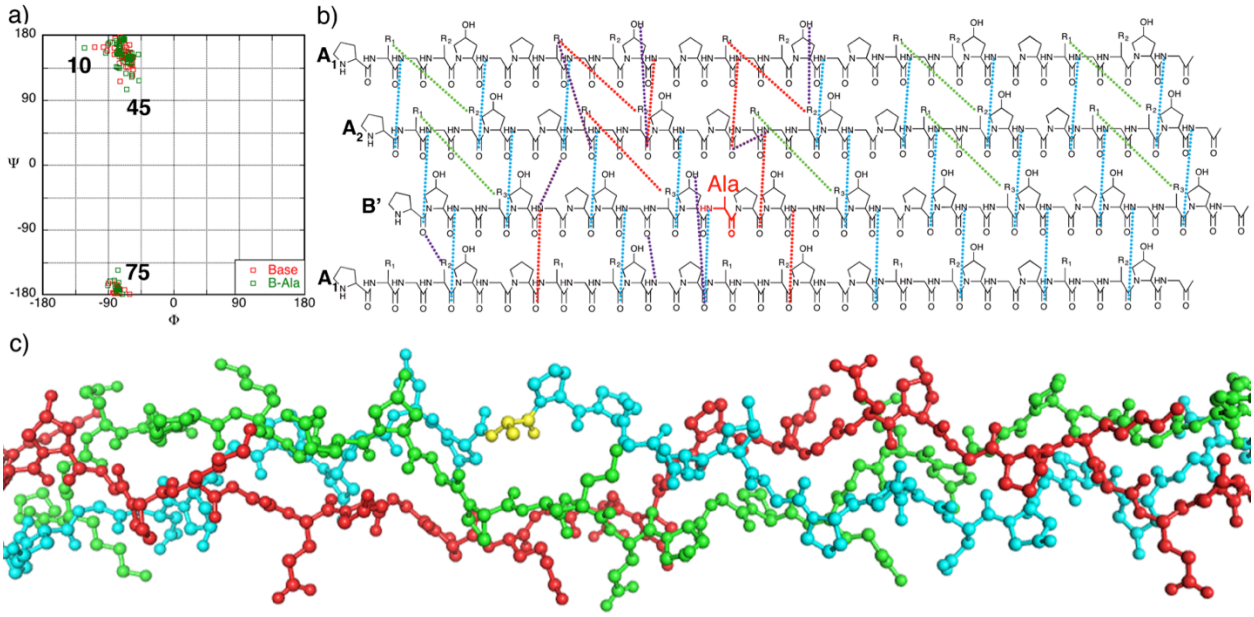


Figure 3.6. a.) Ramachandran plot of B-Ala helix compared to Base b.) Chart of interactions lost and gained in B-Ala model. R_1 represents lysine side chain, R_2 represents glutamate side chain, and R_3 represents aspartate side chain. Blue interactions are expected backbone hydrogen bonds that formed, green are expected axial charge pairs that are formed, red are broken hydrogen bonds or charge pairs, and purple are new compensating interactions c) Generated model of B-Ala. Alanine is shown in yellow.

ppm values, it was possible to determine NOE correlations of neighboring protons for each isotopic label. From these $^{1}H,^{15}N$-NOESY-HSQC (Figure 3.5)spectra restraints were generated and used to guide modeling by Rosetta (constraints available in Appendix 1).

In these models it is apparent that B-Ala compensates for the mutation by a moderate loosening of triple helical topology (Figure 3.6). The helix has a small area of relaxed folding immediately surrounding the mutation site to accommodate the additional methyl group. This results in significant splaying away from the central triple helical axis, which peaks at the mutation site (Figure 3.8). Alterations to the dihedral angles of the peptides prevent three axial charge pairs from forming adjacent to the mutation site, but also enable the helix to quickly re-nucleate and keep the disorder localized. In

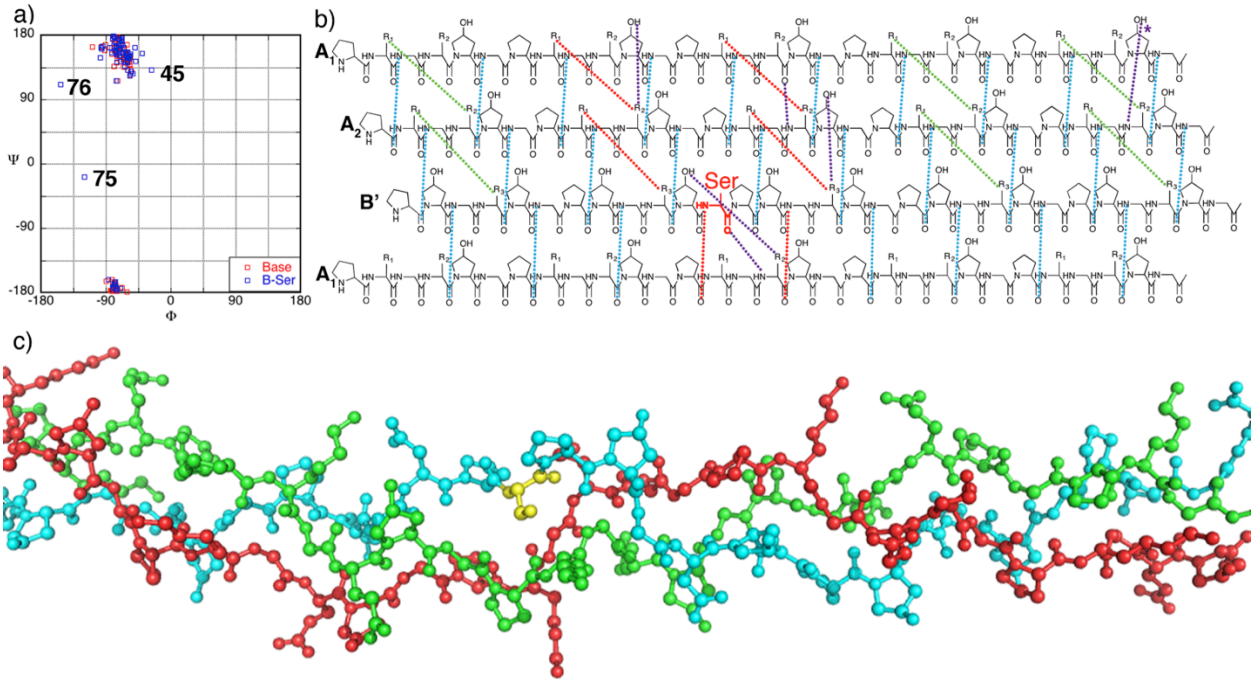


Figure 3.7. a.) Ramachandran plot of B-Ser helix compared to base b.) Chart of interactions lost and gained in B-Ser model. R_1 represents lysine side chain, R_2 represents glutamate side chain, and R_3 represents aspartate side chain. Blue interactions are expected backbone hydrogen bonds that formed, green are expected axial charge pairs that formed, red are broken hydrogen bonds or charge pairs, and purple are compensating interactions that formed. * denotes an interaction far from labeling site that is most likely a product of termini affects, not the mutation. c) Model of B-Ser. Serine shown in yellow

addition to the unsatisfied charge pairs, five interchain hydrogen bonds are lost. To offset these lost interactions, the model suggests that eight new hydrogen bonds form. Three involve hydroxyproline side chains, two involve charged side chains, and the last two are between backbone atoms. Despite the compensating interactions, the mutant structure is less stable as indicated by CD and less well-ordered as indicated by loss of key NOE correlations, and the decrease in ^{15}N-relaxation values.

In spite of being a larger amino acid, B-Ser has fewer interruptions to the backbone hydrogen-bonding network than the alanine mutation, though the dihedral angles of residues surrounding the mutation site deviate to a greater extent than the B-Ala model (Figure 3.7). These altered angles allow the triple helix to compensate for the

disruption, and reduce the local unfolding. This difference from the B-Ala model could in part be attributed to the polar nature of the serine side chain, which could allow for polar interaction with the backbone of other chains, or other amino acid side chains. These interactions are seen in a homotrimeric system[12] and theoretical models.[27,28] The polar group on the mutated amino acid enables the side chain to remain alongside the helix rather than away and allows most normal backbone hydrogen bonds to form. Specifically, the B-Ser triple helix loses four backbone hydrogen bonds, and four stabilizing axial charge pairs. To counterbalance these lost interactions, the model suggests six additional non-canonical hydrogen bonds may form (Figure 3.7, Table 3.4). These new hydrogen bonds are all interchain and present in two distinct types: hydroxyproline OH mediated hydrogen bonds (four) and new backbone hydrogen bonds (two). However, the additional serine polar side chain interactions allow the helix to fold into a closer approximation of the base helix where the radius only deviates to 3.7 Å instead of almost 5 Å as in the B-Ala case (Figure 3.8).

3.4 Comparisons to Other OI Models

When compared to prior studies of *OI*-type mutations, our results suggest these disruptions have limited additive effect. Though prior homotrimeric systems include three mutant strands rather than a single mutant chain, they do not observe significantly increased destabilization due to the triple mutation.[11] Similar to prior reports we observe the relative impact of the amino acids range from the smallest amino acids alanine and serine, which are seen to disrupt the triple helix the least, followed by the larger branched or charged amino acids valine, asparate and arginine which more dramatically disrupt folding. However, the degree of structural disruption per mutations was quite different

between this study and the previous literature where here we observe similar disruption for a single substitution as compared to triple-substituted homotrimers. For example, in published homotrimeric cases there are similar disruptions to the hydrogen-bonding pattern when compared to this study with only one mutated amino acid.[11] This further suggests that the disruption of these OI-type mutations is not an additive effect.

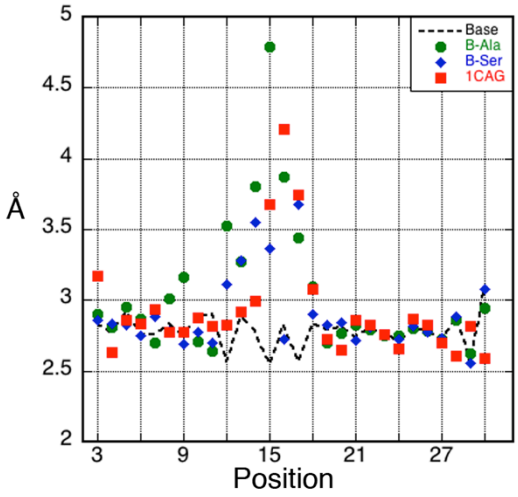


Figure 3.8. Triple Helical Radius of the mutated, base, and 1CAG helices

Additionally, our charge-rich heterotrimer allows our systems to fully re-nucleate past the mutation site, allowing visualization of how natural systems may accommodate these local disruptions. Furthermore, homotrimeric models will not be able to account for all factors influencing OI phenotypes like heterotrimeric mimics, including the presentation of mutants in collagen-binding motifs and the influence of neighboring amino acids from both chains.

Osteogenesis Imperfecta mutations follow a clear pattern based upon the size, charge, and hydrophobicity of the substituted residue. These patterns are replicated in our experiments. The models created here are the first collagen mimetic peptides that fold into the correct composition with an OI-type mutation. Each mutation has a different way

of compensating for the steric effects of the substitution. B-Ala flips the side chain of alanine out of the helix; B-Ser compensates by having a few amino acids with extreme dihedral angles to maintain the hydrogen bonds. From our experiments, all the mutations have an asymmetric effect on the folding of the triple helix, with the greatest disruption on the N-terminal side of the Gly substitution. While these systems compensate with structural disruptions, B-Val, B-Asp, and B-Arg compensate by not forming the intended structure, instead excluding the mutated B-chain and forming the AAA homotrimer.

3.5 Conclusions

Our system illustrates the possibilities for systems to self-assemble into the correct composition and register while still having highly disrupted local structures. This reinforces the robustness of the design principles, and the flexibility of the axial charge pair in heterotrimer collagen mimetic peptide design. These systems are limited, however, by the equilibrium conditions at which the heterotrimer folds. If the energetic benefit of triple helical folding is not high enough the ratio between the monomer and folded state can be unfavorable, especially in cases where the designed sequence incorporates severely destabilizing mutations.

These interactions can be used to design a wide variety of disease models or models of the active binding sites in collagens. On-going investigations are working to incorporate natural guest sequences surrounding a mutation into similar register-specific AAB-type triple helices. These systems will more closely mimic clinical phenotypes and can also examine the effect of rigidity vs. flexibility of the guest region, which may exacerbate or mitigate the disruption of the mutation.[29,30] Several binding sites, including the integrin-binding site FOGER, are known to have OI mutations that may interfere with

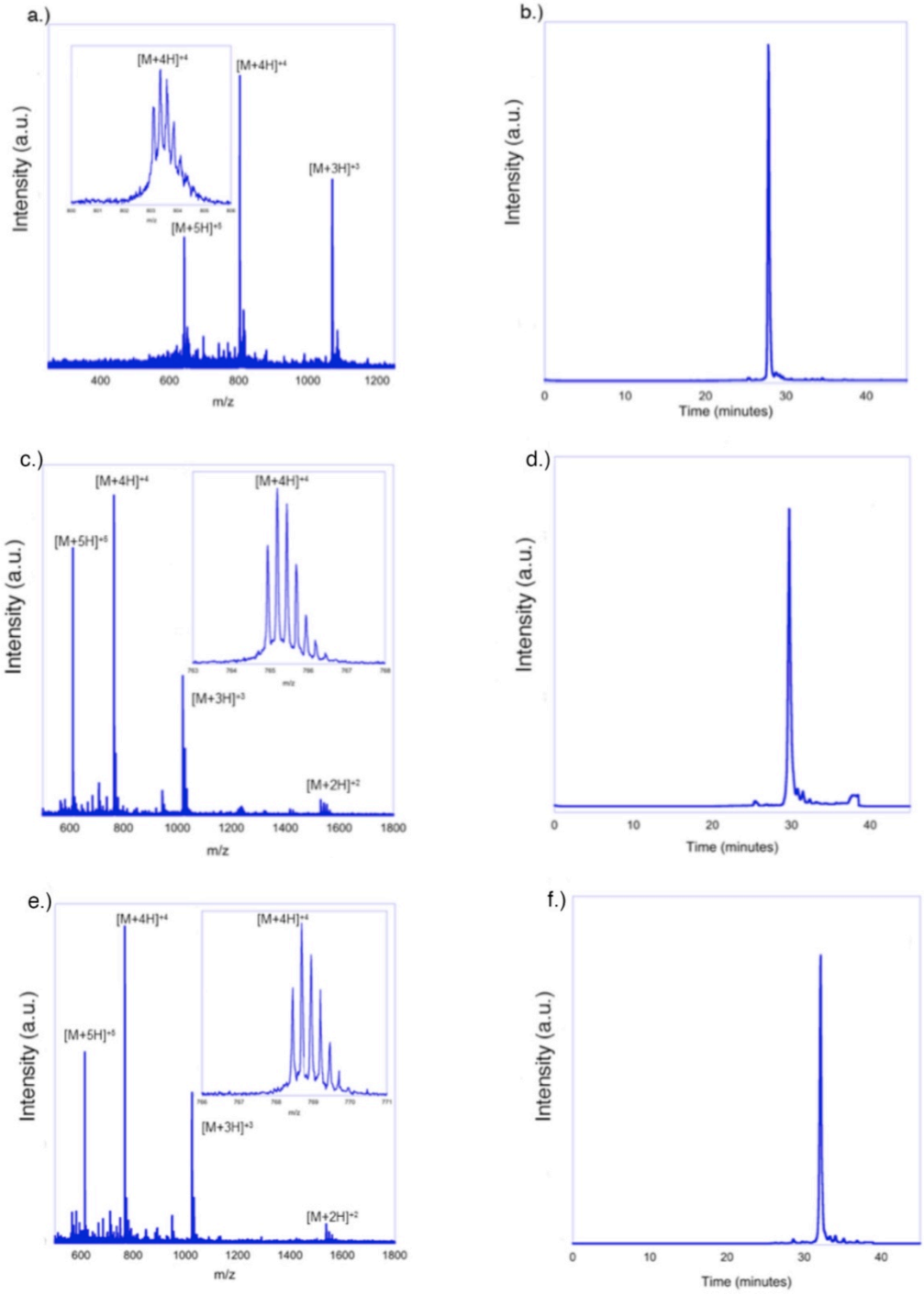


Figure 3.9. a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3210.55 Observed =3211.48) and b) HPLC trace of the Base A peptide. c) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3062.23 Observed =3061.60) and d) HPLC trace of the Base B peptide. e) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3079.25 Observed =3078.08) and f) HPLC trace of the B-Ala B peptide.

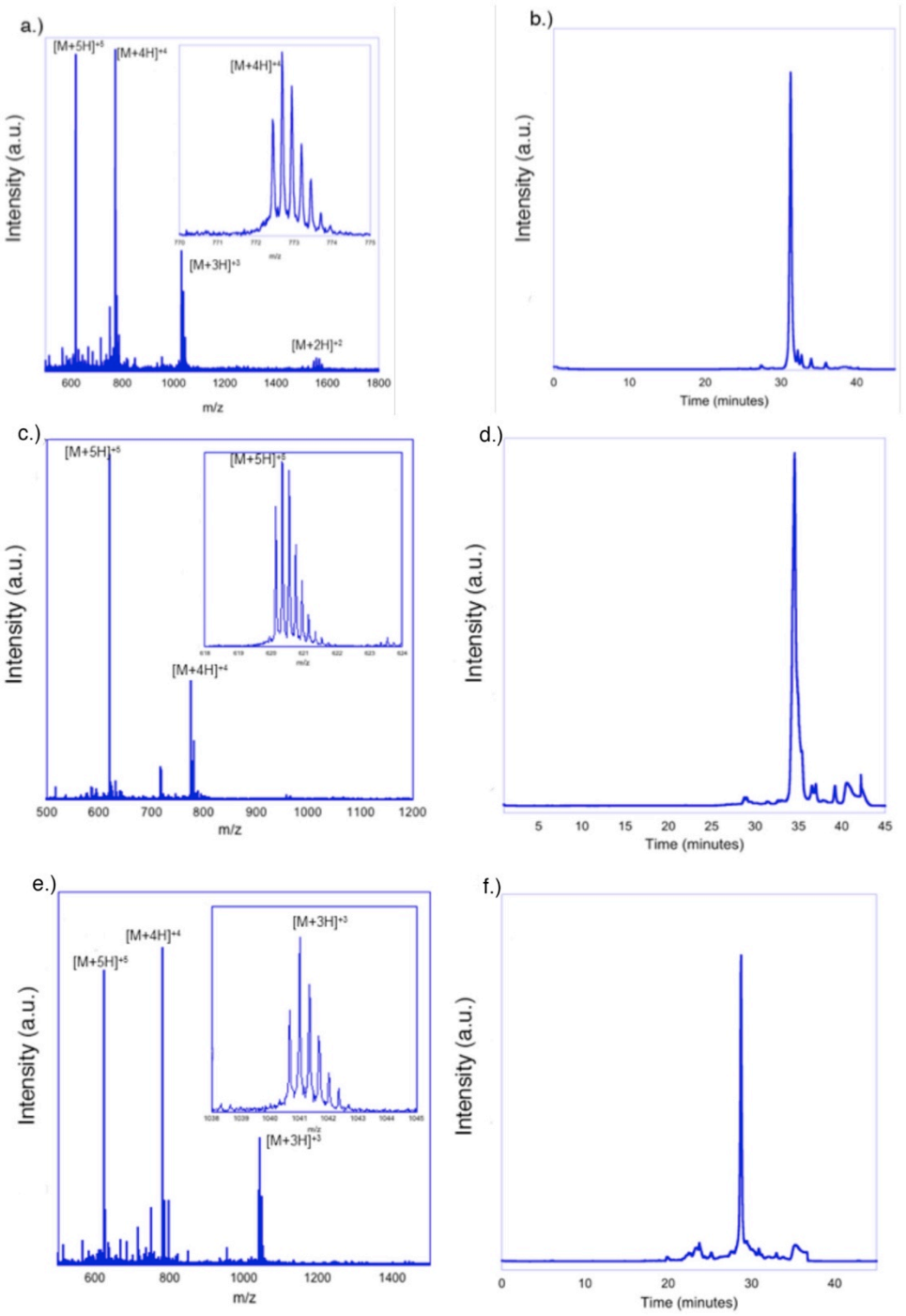


Figure 3.10. a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3095.23 Observed =3094.48) and b) HPLC trace of the B-Ser B peptide. c) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3107.12 Observed =3108.11) and d) HPLC trace of the crude B-Val B peptide. e) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3123.08 Observed =3122.80) and f) HPLC trace of the crude B-Asp B peptide

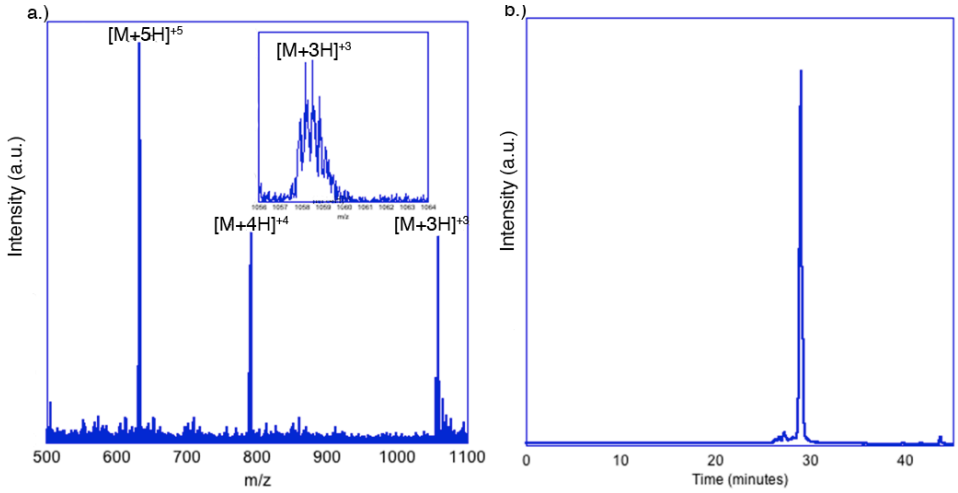


Figure 3.11. a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^+$ =3164.10 Observed =3164.30) and b) HPLC trace of the crude B-Arg B peptide.

binding and regulation. In future investigations, incorporation of natural type I collagen guest sequences could be incorporated to better examine how these mutations affect the presented face of the active site and disrupt binding and local structure.

3.6 Experimental

3.6.1 Peptide Synthesis

Peptides were prepared using standard solid phase synthesis utilizing Fmoc-protected amino acids. The solid phase support was a low loading (0.37 mmol/gram) Rink amide MBHA resin to give the C-terminally amidated product. A 25% by volume piperidine solution in dimethylformamide (DMF) was used for Fmoc-deprotection cycles, and coupling cycles were performed with 2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate methanaminium (HATU) and diisopropylethylamine (DiEA). For cycles that incorporated an ^{15}N isotopic label, reagents were combined in the ratio 1:2:2:3 (resin: amino acid: HATU: DiEA); for all other cycles the ratio used was 1:4:4:6. Acetylation at the N-terminus was completed with a 1:8:4 mixture of resin: acetic anhydride: DiEA in dichloromethane; this step was

performed twice. Cleavage from the resin was performed for three hours using 1,2-ethanedithiol: H_2O: triisopropysilane (1:1:0.4) in 20 mL trifluoroacetic acid (TFA). The reaction mixture was then drained and the TFA removed by rotary evaporation under reduced pressure. When less than 5 mL solution remained, ice-cold diethyl ether was used to precipitate out the crude peptide. The peptide was centrifuged to obtain a pellet; this step was repeated two times using 100 mL of diethyl ether to further remove TFA.

3.6.2 Peptide Purification

After drying overnight the crude peptides were dissolved in deionized water and purified by reverse phase high-pressure liquid chromatography (HPLC). Peptide molecular weights were confirmed with electrospray ionization time of flight mass spectrometry (ESI TOF-MS) (Figures 3.9-11) and lyophilized to give the purified peptide

3.6.3 Sample Preparation

The lyophilized powders were dissolved in autoclaved deionized water to prepare stock solutions. From these solutions NMR samples were prepared by dilution to a final concentration of 3 mM in 9:1 H_2O:D_2O in a 10 mM PO_4 buffer. Circular dichroism (CD) samples were prepared to a final concentration of 0.3 mM. All samples were annealed for 15 minutes above 80 °C, allowed to cool to room temperature and equilibrated for 1 week at room temperature, followed by 1 week incubation at 4 °C. CD wavelength and melting spectra were used to confirm peptide folding for each sample before NMR samples were measured.

3.6.4 NMR Spectroscopy

All TOCSY, and $^{1}H,^{15}N$-HSQC and $^{1}H,^{15}N$-NOESY-HSQC experiments were performed on a Varian spectrometer of 800 MHz containing a cryogenic probe. The

nitrogen carrier frequency was set at 110 ppm, while the proton carrier frequency set to the water signal. All experiments were performed at 10 °C. Chemical shift assignments of residues were made using TOCSY spectra. For TOSCY spectra 1120 x 512 complex points and 128 scans were acquired; the mixing time was 85 ms and sweep width 12 kHz in both dimensions. The $^{1}H,^{15}N$-HSQC were collected with 1040 x 256 complex points for a total of 16 scans. A sweep width of 10 kHz in the proton and 2432 Hz in the nitrogen dimension was used. To compare relative peak volumes of folded and unfolded states, for each $^{1}H,^{15}N$-HSQC of an AAB' heterotrimer the total peak volume was set to 1. Because of the difference in the relaxation rate of the monomer and folded helix no quantitative relationship between the populations can be determined from this ratio, but a general comparison between the mutation types may be made. A 2D version of the 3D $^{1}H,^{15}N$-NOESY-HSQC was collected with 100 ms mixing time, 64 scans, 1200 x 256 complex points; the spectral window was 12 kHz in the direct dimension and 12 kHz for the indirect dimension. ^{15}N relaxation experiments were performed on 3 mM samples as previously described on a 500 MHz Varian spectrometer.[26] Data was processed and analyzed using NMRdraw, and ccpnmr processing suites.[27, 28]

3.6.5 CD Spectroscopy

Circular dichroism (CD) experiments were collected on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller. Ellipticity measurements were performed between 190 and 250 nm to characterize helical secondary structure of the samples. The maximum wavelength recorded in this region was then selected to be monitored between 5 and 85 °C at a heating rate of 10 °C/hour. The first derivative of this experiment was calculated using a Savitzky-Golay smoothing

algorithm, and the melting temperature was assigned to the minimum(s) detected. The molar residue ellipticity (MRE) value was calculated with the following equation, where Θ represents the experimental ellipticity in millidegrees, m is the molecular weight of the triple helix (g/mol), c is the peptide concentration (milligrams/milliliter), l is the pathlength of the cuvette (cm), and nr represents the total number of residues in the triple helical system.

$$[\theta] = \frac{\theta}{c \times l \times nr \times 10}$$

3.6.6 Molecular Modeling

From NMR studies, specifically $^{1}H,^{15}N$-NOESY-HSQC, constraints were created for the folded heterotrimers (available Appendix 1). Negative constraints were also created for correlations that existed in the base structure, but not the mutated structure. Positive constraints were set from 0-4 Å, and negative from 6-10 Å. These constraints were utilized in the Rosetta Relax application[29,30] to generate a set of structures. Generated structures were assessed for compliance to the constraints. For all positive correlations the distances are below 5 Å, and above 5 Å for negative constraints. Samples were run iteratively with the best structure from the previous run used with changes to the standard deviation until a compliant structure was found. For constraints that were hard to satisfy, the standard deviation was decreased to increase the energetic penalty for longer deviations. The structure with the best compliance to NMR data was selected.

3.7 References

(1) Ramachandran, G. N. G.; Kartha, G. G. *Nature*. **1954**, *174*, 269–270.
(2) Rich, A.; Crick, F. H. C. *J. Mol. Biol*. **1961**, *3*, 483–506.
(3) Jalan, A. A.; Hartgerink, J. D. *Biomacromolecules*. **2013**, *14*, 179–185.

(4) Fallas, J. A.; Lee, M. A.; Jalan, A. A.; Hartgerink, J. D. *J. Am. Chem. Soc*. **2012**, *134*, 1430–1433.

(5) Xu, F.; Zahid, S.; Silva, T.; Nanda, V. *J. Am. Chem. Soc*. **2011**, *133*, 15260–15263.

(6) Jalan, A. A.; Hartgerink, J. D. *Curr. Opin. Chem. Biol*. **2013**, *17*, 960–967.

(7) Kielty, C. M.; Hopkinson, I.; Grant, M. E. Collagen: The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix. In *Connective Tissue and Its Heritable Disorders*, 1; Royce, P. M., Steinmann, B., Eds.; John Wiley & Sons Inc.: Hoboken. **2003**; 103-147.

(8) Marini, J. C.; Forlino, A.; Cabral, W. A.; Barnes, A. M.; San Antonio, J. D.; Milgrom, S.; Hyland, J. C.; Korkko, J.; Prockop, D. J.; De Paepe, A.; Coucke, P.; Symoens, S.; Glorieux, F. H.; Roughley, P. J.; Lund, A. M.; Kuurila-Svahn, K.; Hartikka, H.; Cohn, D. H.; Krakow, D.; Mottes, M.; Schwarze, U.; Chen, D.; Yang, K.; Kuslich, C.; Troendle, J.; Dalgleish, R.; Byers, P. H. *Hum. Mutat*. **2007**, *28*, 209–221.

(9) Byers, P. H. *Philos. Trans. R. Soc., B*. **2001**, *356*, 151–158.

(10) Kuivaniemi, H.; Tromp, G.; Prockop, D. J. *Hum. Mutat*. **1997**, *9*, 300–315.

(11) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science*. **1994**, *266*, 75–81.

(12) Li, Y.; Brodsky, B.; Baum, J. *J. Biol. Chem*. **2009**, *284*, 20660–20667.

(13) Buevich, A.; Baum, J. *Philos. Trans. R. Soc., B*. **2001**, *356*, 159–168.

(14) Xiao, J.; Madhan, B.; Li, Y.; Brodsky, B.; Baum, *Biophys. J*. **2011**, *101*, 449–458.

(15) Baum, J.; Brodsky, B. *Curr. Opin. Struct. Biol*. **1999**, *9*, 122–128.

(16) Hwang, E. S.; Brodsky, B. *J. Biol. Chem*. **2012**, *6*, 4368–4375.

(17) Xiao, J.; Sun, X.; Balaram, M.; Brodsky, B.; Baum, J. *J. Biol. Chem.* **2015**, *290*, 24201-24209.

(18) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *129*, 7509–7515.

(19) Byers, P.H.; Wallis, G. A.; Willing, M. C. *J. Med. Genet*. **1991**, *28*, 433-442.

(20) Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. *Proc. Natl. Acad. Sci.* **2000**, *97*, 4273–4278.

(21) Hyde, T. J.; Bryan, M. A.; Brodsky, B.; Baum, J. *J. Biol. Chem*. **2006**, *281*, 36937–36943.

(22) Mooney, S. D.; Klein, T. E. *Mol. Cell. Proteomics*. **2002**, *1*, 868–875.

(23) Radmer, R. J.; Klein, T. E. *Biochemistry*. **2004**, *43*, 5314–5323.

(24) Bryan, M. A.; Cheng, H.; Brodsky, B. *Biopolymers*. **2011**, *96*, 4–13.

(25) Xiao, J.; Yang, Z.; Sun, X.; Addabbo, R.; Baum, J. *J. Struct. Biol*. **2015**, *192*, 127-137

(26) Acevedo-Jake, A. M.; Jalan, A. A.; Hartgerink, J. D. *Biomacromolecules*. **2014**, *16*, 145–155.

(27) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomole. NMR*. **1995**, *6*, 277–293.

(28) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins*. **2005**, *59*, 687–696.

(29) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. *Methods Enzymol.* **2004**; 383, 66–93.

(30) Leaver-Fay, A.; O'Meara, M. J.; Tyka, M.; Jacak, R.; Song, Y.; Kellogg, E. H.; Thompson, J.; Davis, I. W.; Pache, R. A.; Lyskov, S.; Gray, J. J.; Kortemme, T.; Richardson, J. S.; Havranek, J. J.; Snoeyink, J.; Baker, D.; Kuhlman, B. *Methods Enzymol*. **2013**, *523*, 109–143.

Chapter 4: Comparison of *Osteogenesis Imperfecta* Mutations in the A, B, and all Three Chains*

4.1 Introduction

A collagen triple helix is a supramolecular complex that consists of three peptide chains in the polyproline II conformation[1, 2] that self-assemble to create a right-handed superhelix. Each chain has a three amino acid repeat of Xaa-Yaa-Gly with stabilizing hydrogen bonds that form between a glycine NH and the backbone carbonyl of an Xaa position in an adjacent chain. In a triple helix the separate peptide strands may be the same or different sequences which determines the composition of the triple helix. When all three peptides are identical the helix is called an AAA homotrimer, when there are two unique peptides the helix is an AAB heterotrimer and when all three are distinct the helix is called an ABC heterotrimer. Because there is a one amino acid offset between each chain of the triple helix, the order of the peptides can create unique triple helical registers in heterotrimers. For example, an AAB triple helix will be distinct from an ABA because different chains will be in the leading, middle, and lagging positions.

The design of composition and register specific collagen triple helices has advanced allowing for more complex models of disease conditions, specifically the mutations found in *Osteogenesis Imperfecta* (OI).[3] OI is a disease that usually results from mutations of glycines in the Xaa-Yaa-Gly repeat in type I collagen. These mutations cause a wide range of disease phenotypes based upon the identity of the amino acid

---

* The work in this chapter was published as Clements, K.A.; Acevedo-Jake, A.M.; Walker, D. R.; Hartgerink, J.D. "Glycine Substitutions in Collagen Heterotrimers Alter Triple Helical Assembly" *Biomacromolecules*. **2017**, *18*, 617-624.

AAJ and KAC synthesized and purified the peptides. AAJ performed NMR experiments. KAC performed CD and modeling experiments. DRW did helical twist analysis. DRW, AAJ, KAC, and JDH analyzed the data and wrote the manuscript.

replacing glycine and the location of the mutation in the sequence.[4, 5] From clinical data, experiments, and theoretical models the severity of the disease has been correlated with the size and charge of the substituted amino acid, specifically Ala < Ser < Cys < Arg < Glu < Asp < Val in order from least to most disruptive.[4, 6-8] The placement of the mutation in the A chain or B chain also changes the severity because, in principle, two mutations are incorporated when the A chain is mutated, and only one when the B chain is mutated.[4, 5]

Before register and composition control were available, the first OI models were homotrimers. Though these systems gave significant insight into glycine mutations in collagen, they cannot accurately account for structural and thermal disruptions found in natural heterotrimeric mutations. For a glycine to alanine substitution in a homotrimer the mutation is always incorporated into all three chains of the triple helix.[9] Amongst other changes to the helix structure, this has been shown to result in the loss of four intermolecular backbone hydrogen bonds. This local disruption decreases the melting temperature, T_m, of the helix by 29 °C compared to the base $(POG)_{10}$ sequence. Further mutations have been incorporated into a homotrimeric triple helix and studied with circular dichroism and NMR.[9-11] In homotrimers generally, a mutation typically results in the loss of three to four backbone hydrogen bonds and a sharp decrease in thermal stability. With the development of ABC heterotrimeric triple helices with compositional control, mutations could selectively be made in just one or two chains. Previously, our lab prepared an ABC triple helix with a central AAB guest region to model mutations in one, two and all three chains.[12] We saw that once the triple helix was mutated, successive mutations had a smaller impact on the stability of the helix; specifically the

destabilization was not linear with respect to the number of mutations. However, this system gave no information about composition or register, and also lacked structural characterization of the mutation site.

Recently, utilizing a robust design of an AAB register and composition controlled triple helix, just the B chain of the triple helix was mutated to study the impact on the triple helix stability and structure.[13] This system showed a clear correlation between the structural disruption and the size or charge of the substituted amino acid. Smaller uncharged amino acids exhibited less disruption to the triple helix than large charged amino acids. The mutations also showed an impact on the composition and register, where increasing size or charge of the mutated amino acid led to disruptions away from the desired AAB' population. Here we complement that study by analyzing the impact of mutations to the A chain of AAB heterotrimers. We also examine systems in which both the A and B chains have been mutated. Our analysis shows the strong driving force for triple helices to avoid double mutations and the changes in helical composition that arise to accomplish this despite the loss of many stabilizing charge-pair interactions.

4.1.1 Peptide Design

To mimic natural OI disruptions, we used a highly stable AAB register controlled system as our base sequence and introduced glycine mutations.[14] In order to explore the molecular structure of the triple helices, ^{15}N isotopically labeled glycines are placed around the mutated region and also were placed N-terminally (Fig. 4.1) to give information on refolding. Additionally two different amino acid substitutions (alanine or serine) were tested. Here, systems with a mutation only in the A chain are called A-Ala or

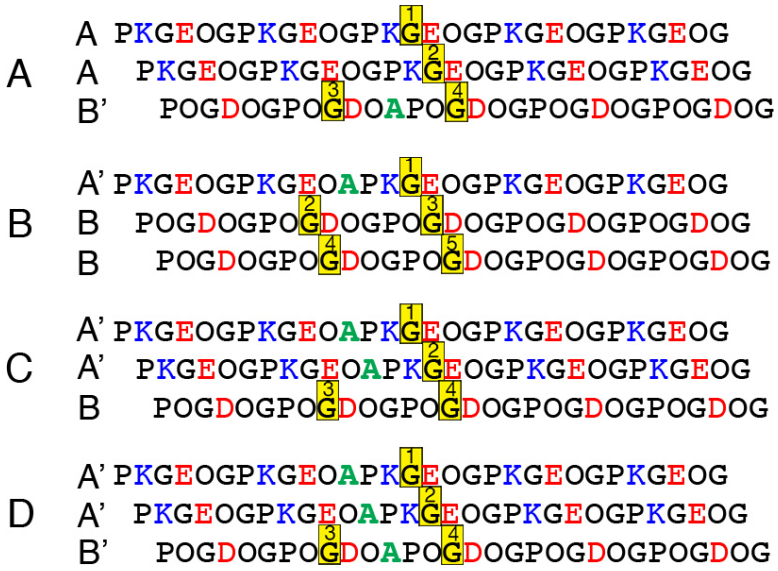


Figure 4.1. Sequences, compositions and registers of peptides used. Red amino acids are negatively charged, blue are positively charged. The yellow boxed glycines are ^{15}N isotopically labeled. A is the B-Ala mutant, B is A-Ala mutant which is observed to form, C is A-Ala mutant which was intended but is not observed to fold, D is the 3-Ala mutant. The serine mutants are identical other than the substitution of alanine for serine. A-Ser depending on the amino acid substituted. The double mutant systems are called 3-Ala and 3-Ser.

4.2 Characterization of Mutant Triple Helices

4.2.1 Base AAB Triple Helix

To improve upon the physiological relevance of OI models and gain better insight into natural systems, a previously designed register and composition controlled AAB heterotrimer was used to mimic the composition of type I collagen.[14] To ease composition verification in our systems a single ^{15}N-glycine label is included in the A chain and two ^{15}N-glycine labels in the B chain. This produces four folded ^{1}H,^{15}N-HSQC peaks in an AAB triple helix as well as a peaks each for the *cis* and *trans* conformations of the labels in unfolded single A or B chains.[13] This is a well-folded system, shown by the high ratio of trimers to monomers, and the relatively high (> 8) relaxation values of the ^{1}H,^{15}N-HSQC peaks indicating good local stability. R1/R2 relaxation values below two typically

Substitution	*Name*	*Heterotrimer T_m*	*Folding*	*R2/R1 of Heterotrimer*	*Monomer Rel Vol.*	*AAB' Rel. Vol.*	*A'BB Rel. Vol.*
Glycine[14]	*Base*	*51.0 °C*	*Single composition, single register AAB*	*14.9, 9.6, 21.3, 8.0*	*0.33*	*0.67*	-
Alanine[13]	*B-Ala*	*30.5 °C*	*Single composition, single register AAB'*	*20.8, 10.7, 13.8, 10.0*	*0.44*	*0.56*	-
Serine[13]	*B-Ser*	*29.5 °C*	*AAB' folded as well as AAA homotrimer*	*15.7, 13.5, 18.6, 17.8*	*0.61*	*0.37*	-
Alanine	*A-Ala*	*34.0 °C*	*Single composition, single register A'BB*	*7.7, 8.9, 11.9, 7.8, 6.0*	*0.50*	-	*0.50*
Serine	*A-Ser*	*30.0 °C*	*Single composition, single register A'BB*	*8.6, 11.1, 6.7, 10.4, 9.4*	*0.82*	-	*0.18*
Alanine	*3-Ala*	*19.0 °C*	*Single composition A'A'B'*	*5.3, 2.4, 3.0, 4.5*	*0.81*	*0.19*	-
Serine	*3-Ser*	*18.0 °C*	*Single composition, single register A'A'B'*	*8.3, 4.9, 5.3, 5.8*	*0.83*	*0.17*	-

Table 4.1. NMR Experimental Results of the different Osteogenesis Imperfecta model systems. Heterotrimer T_m is the melting point of the folded heterotrimer by CD. The R2/R1 value, or relaxation value, is for each isotopic label (G1-G5) in the order from leading to lagging chain. Relative volumes are the relative volume of the peaks when the total peak volume is set to 1 (see experimental procedures for more details).

indicate monomeric peptides while higher values indicate increased local rigidity of a folded helix. The CD melting spectra corroborates the robustness of the base heterotrimer showing a single transition at 51 °C. The helical twist of the base triple helix has a typical amount of variation and falls mostly within the range of a typical POG triple helix. It does have one point outside of the range, which is believed to be the result of strain induced by the surrounding charge pairs.

4.2.2 A-Chain Mutants

Like the base triple helix, the triple helices A-Ala and A-Ser are mixed at a 2:1 ratio of A':B and are designed to fold into an A'A'B triple helix (A' indicates the mutated version of the A peptide) composition and register. As seen for the base triple helix, this composition should yield four peaks in the trimer region of the $^{1}H,^{15}N$-HSQC spectrum corresponding to the four different isotopic labels in the trimer (Figs. 4.1-2). Instead however, the NMR showed five distinct trimer signals of equal volume for both

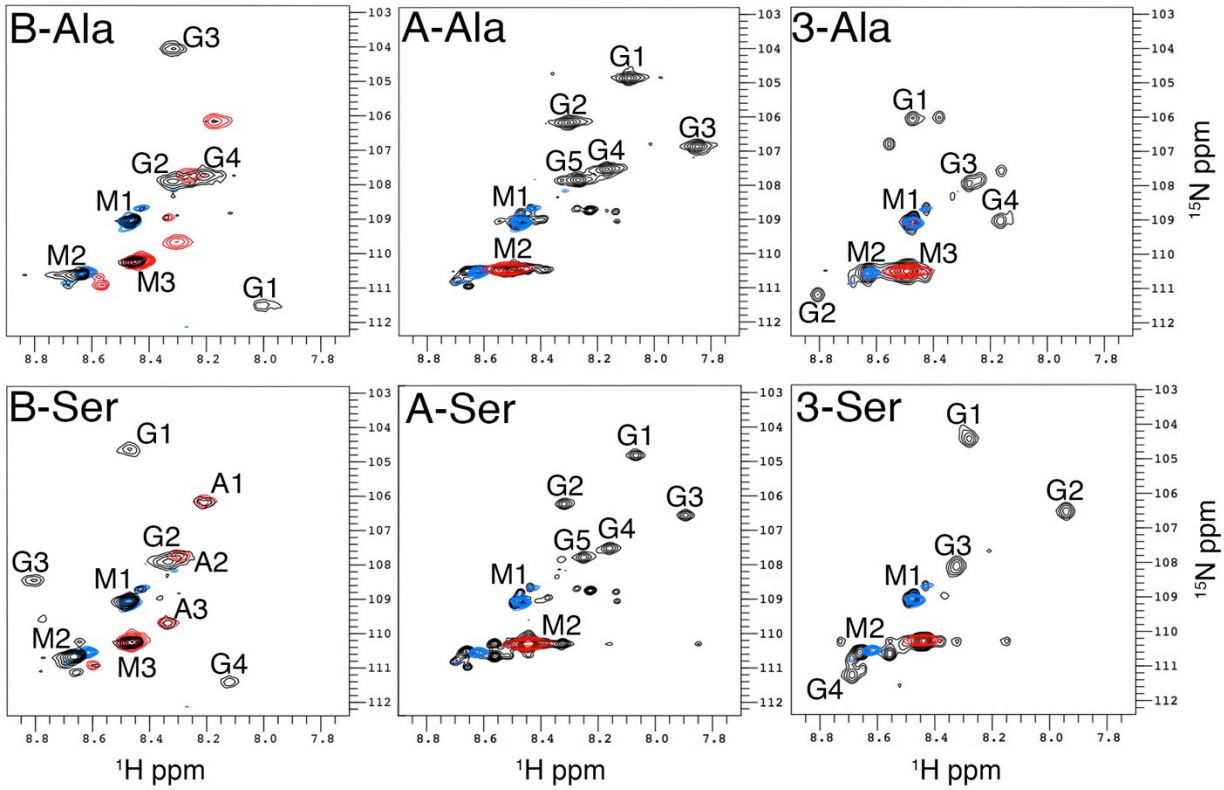


Figure 4.2. ^{1}H,^{15}N-HSQC of the peptide systems. Black spectra are the AAB mixture. Blue is only the B strand used in the system. Red is only the A strand used in the system. G1-G5 denote the heterotrimer present and are labeled in the same order as shown in Figure 1. A1-A3 denote the AAA homotrimers. M1-M3 denote the peptide monomers. B-Ala and B-Ser spectra are reproduced from reference 13 for comparison.

the A-Ala and A-Ser cases, labeled G1-G5 in Fig. 4.2. In our labeling scheme, these five peaks can only appear if the composition has been altered from the intended A'A'B composition to an A'BB composition. One peak represents the label in the A' chain, and the B chains each contain two ^{15}N labels. (Fig. 4.2). While type I collagen is quite long (~1000 residues) our shorter 30-mers are not able to compensate for the local disruption that would be induced by two glycine mutations in the A chain, and instead preferentially form the A'BB composition. Our results suggest that the energetic barrier needed to compensate for two A chain mutants outweighs the stabilization that the A'A'B system could produce by folding; in this manner natural systems may also be affected and suffer from poor folding equilibria.

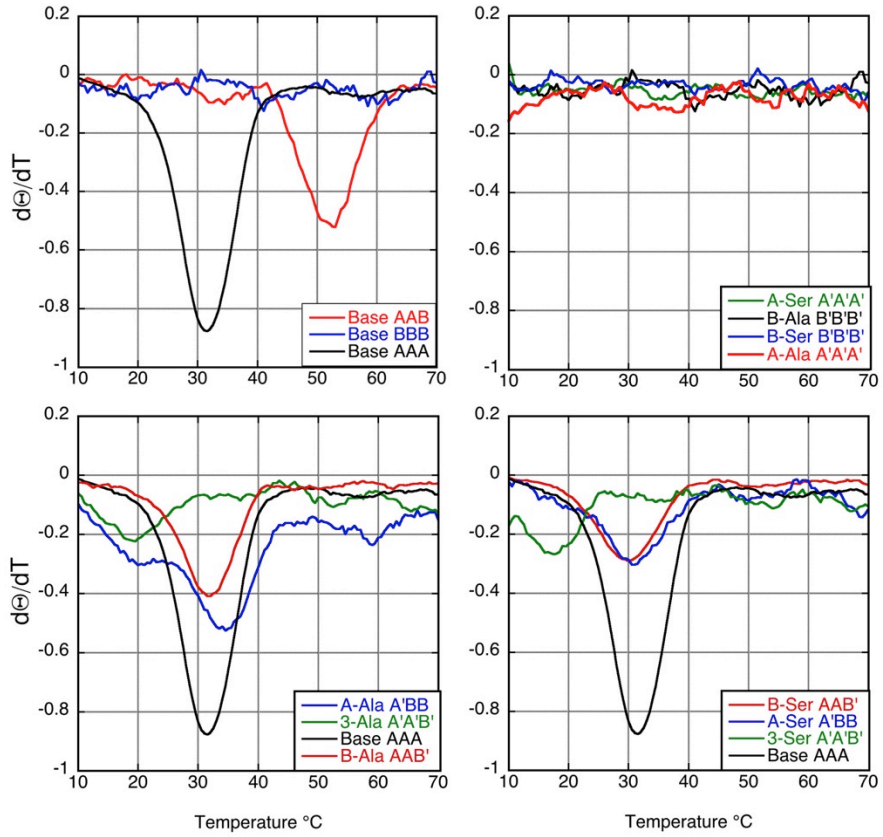


Figure 4.3. Circular Dichroism. First derivatives of the melting plots are shown for clarity. The melting temperature is defined as the minimum of the curve. The Base AAA peptide is shown in all three panels because it is a competing homotrimer in the solution.

A-Ala melted at 34 °C, making it less stable than the base T_m of 51 °C, but more stable than a previously reported B-Ala triple helix which melted at 30.5 °C (Table 4.1, Fig. 4.3). Both A-Ala and B-Ala have disrupted equilibria, as seen by the higher ratio of monomer to heterotrimer than in the base system. However, the A-Ala A'BB helix that folds has five highly stabilizing charge pairs while the B-Ala AAB' helix has seven that form (Table 4.2). The increase in melting point is in contrast with the relaxation values (R2/R1) for the isotopic labels (7.7, 8.9, 11.9, 7.8, and 6.0) (relaxation values are in order from leading to lagging chain) which are also lower than both the base system and B-Ala, indicating A-Ala is a less ordered helix and is likely to have a more disordered conformation around the mutation (Table 4.1). Commonly, relaxation values are

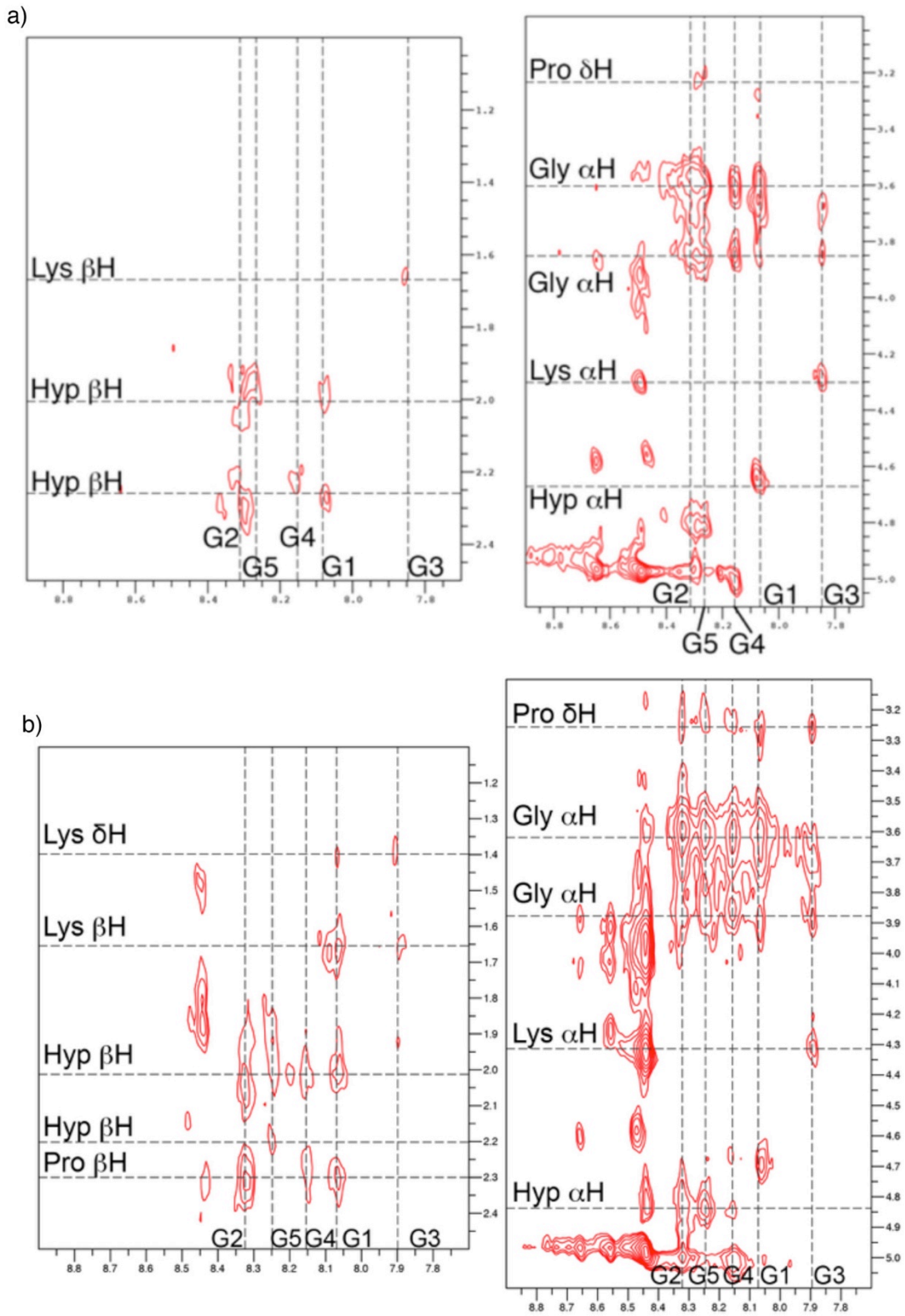


Figure 4.4. NOESY-HSQC of A-Ala (a) and A-Ser (b).

used to measure the quality of triple helical folding, and it is intuitive to assume that a well-folded helix would have a higher melting point than one with lower relaxation values. However, we see the opposite, where our less well-folded helix A-Ala has a higher melting point than our better-folded helix, B-Ala. This suggests that the melting point is not comprehensive in accessing the quality of a triple helix, and other methods such as NMR and molecular modeling must be used.

A model was created of the A-Ala A'BB heterotrimer utilizing constraints from NMR (Fig. 4.4). When compared the B-Ala AAB' model, the A-Ala A'BB model has four more formed backbone hydrogen bonds, but two less formed axial charge pairs as shown in Fig. 4.5 and Table 4.2. This again is surprising because, the more disordered helix, according to the relaxation values, has more backbone hydrogen bonds formed. However, looking at the helical twist of the mutated triple helices gives additional insight to this issue. The helical twist is the angle (in degrees) that the triple helix rotates from one cross section to the next (see supporting information and experimental procedures for a detailed description). For a general $(POG)_{10}$ triple helix the helical twist varies between approximately 95-115 ° from one cross-section to the next. In the B-Ala case (Fig. 4.7) the helical twist varies more dramatically with values from 70 to 130 ° with many swings from smaller to much larger angles. This shows a helix with a tightly twisted cross-sections (larger angle) adjacent to loosely twisted (smaller angle) cross-sections. However, while the A-Ala case has variation, it is primarily to a larger angles near the mutation site. We propose that the tightly twisted region around the mutation site allows the triple helix to compensate for the steric effects of the mutation, while the B-Ala case is more constrained by formed axial charge pairs. These charge pairs force the variation

Triple Helix	*Lost hydrogen bonds*	*Lost charge pairs*	*Compensating Interactions*	*Formed charge pairs*
B-Ala^{13}	*5*	*3*	*8*	*7*
B-Ser^{13}	*2*	*4*	*6*	*6*
A-Ala	*1*	*0*	*0*	*5*
A-Ser	*1*	*0*	*5*	*5*
3-Ala	*1*	*1*	*3*	*9*
3-Ser	*2*	*1*	*7*	*9*

Table 4.2. Results of the modeling experiments. Lost hydrogen bonds are backbone hydrogen bonds that are expected to be found in the model, but are not present. Lost charge pairs are designed axial charge pairs that are not formed in the model. Compensating interactions are hydrogen bonds that are not present in the base structure, but are formed in the model. Formed charge pairs are designed axial charge pairs that are present in the model.

in helical twist, resulting in a triple helix with more interactions, but which is more constrained and relatively inflexible. These systems show the demands of charge pair formation to the triple helix, as well as the flexibility needed to compensate for a glycine mutation.

A-Ser melted at 30 °C, similar to the B-Ser triple helix that melted at 29.5 °C, and B-Ala at 30.5 °C. The similar melting point is surprising because in the models B-Ala and B-Ser have formed seven and six highly stabilizing charge pairs, respectively, while B-Ser has only formed five. The relaxation values of the ^{1}H,^{15}N-HSQC peaks the in the A-Ser spectrum were similar to the A-Ala cases, suggesting a moderately disordered helix. These relaxation values are lower than the B-Ala and B-Ser, suggesting a less ordered helix with a similar melting point. When comparing the A-Ser and A-Ala spectra, much less A-Ser was seen to fold by NMR with a relative volume of 0.18 trimer, suggesting this mutation affects not only the structure and composition of the assembled peptide, but also negatively influences its folding equilibria (Table 4.1).

A-Ser was also modeled using constraints from NMR. The A-Ser A'BB helix was similar to the A-Ala A'BB helix with one lost hydrogen bond and five axial charge pairs

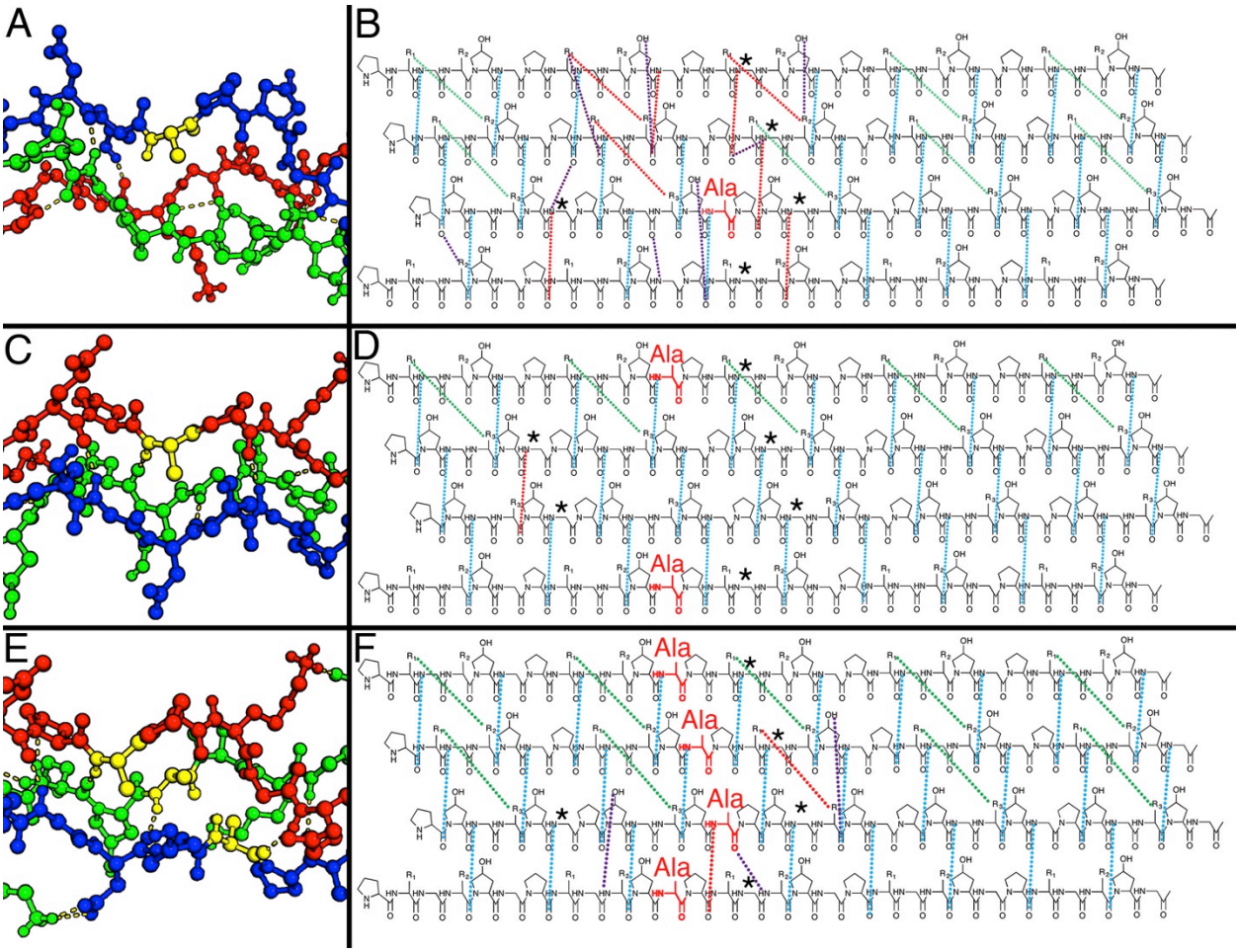


Figure 4.5. Interactions in (a,b) B-Ala,[13] (c,d) A-Ala, (e,f) 3-Ala. a, c, and e are models of the mutation site . Yellow amino acids are the mutated alanines. The red chain is the leading strand, the green chain is the middle strand, and the blue chain is the lagging strand in the triple helix. Hydrogen bonds are shown in the yellow dashed lines. b, d, and f are schemes of the interactions found in the models. Blue dotted lines are expected backbone hydrogen bonds that are formed in the final model. Green dotted lines are expected charge pairs that are formed in the final model. Red dotted lines show expected interactions that are not formed in the final model. Purple dotted lines are compensating interactions that appear in the final models that do not form in the un-mutated model. Asterisks denote the isotopically labeled glycines. R_1 denotes a lysine side chain, R_2 denotes a glutamic acid side chain, R_3 denotes an aspartic acid side chain.

formed (Table 4.2). However, A-Ser had additional hydrogen bonds formed that were not present in A-Ala, three of which were between the serine OH side chain and the peptide

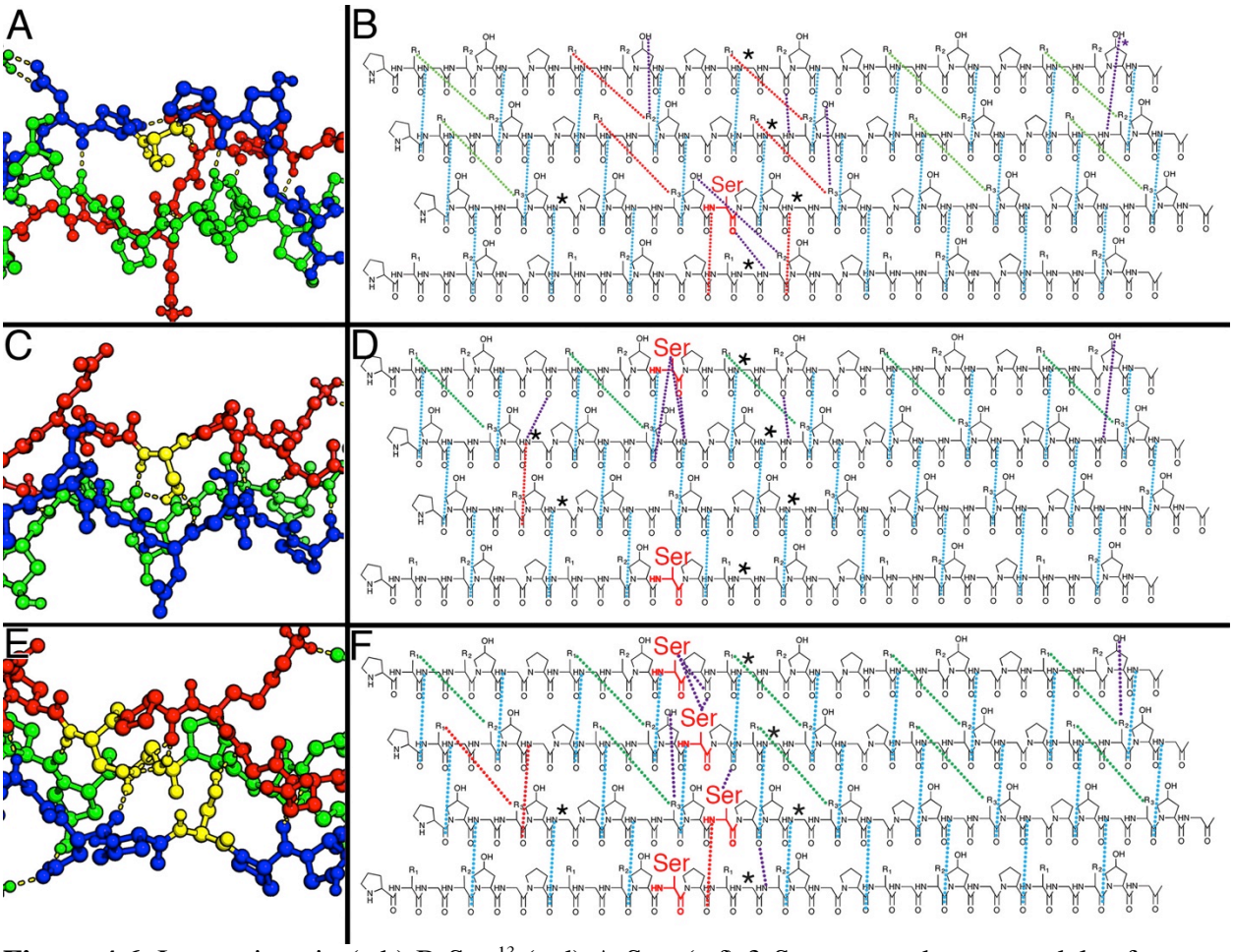


Figure 4.6. Interactions in (a,b) B-Ser,[13] (c,d) A-Ser, (e,f) 3-Ser. a,c, and e are models of the mutation site . Yellow amino acids are the mutated serines. The red chain is the leading strand, the green chain is the middle strand, and the blue chain is the lagging strand in the triple helix. Hydrogen bonds are shown in the yellow dashed lines. B, d, and f are plots of the interactions found in the models. Blue dotted lines are expected backbone hydrogen bonds that are formed in the final model. Green dotted lines are expected charge pairs that are formed in the final model. Red dotted lines show expected interactions that are not formed in the final model. Purple dotted lines are compensating interactions that appear in the final models that do not form in the un-mutated model. Asterisks denote the isotopically labeled glycines. R_1 denotes a lysine side chain, R_2 denotes a glutamic acid side chain, R_3 denotes an aspartic acid side chain.

backbone (Fig. 4.6). These additional interactions stabilize the triple helix, but also constrain the area around the mutation. This can be seen when comparing the A-Ser and A-Ala helical twist, where the unconstrained A-Ala has one increase to 130° and quickly

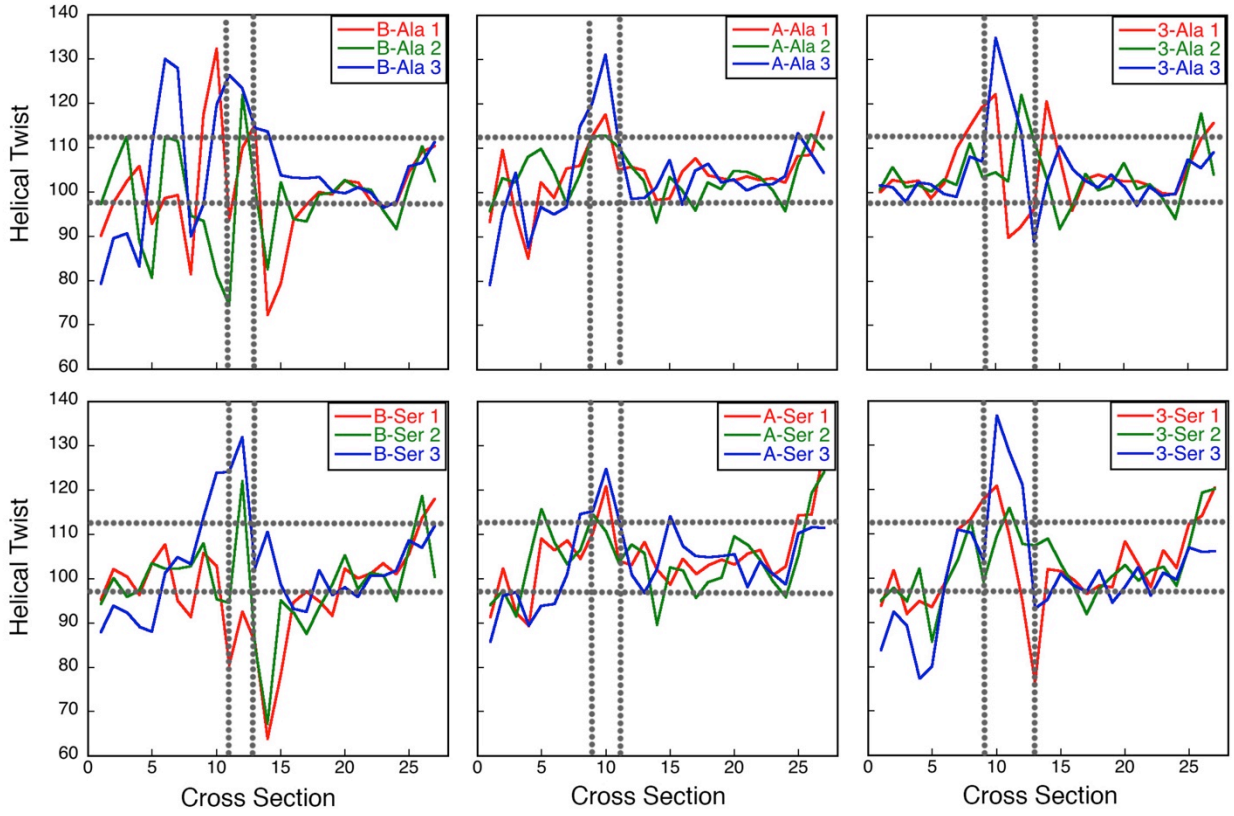


Figure 4.7. Helical twist of the peptide systems. The first (leading) chain of the triple helix is in red, the second (middle) is in green, and the third (lagging) is in blue. The vertical grey area is the mutated region, and the horizontal grey area is the range of helical twist in a POG model. The x-axis is the cross-section of the triple helix, starting where the lagging strand joins the triple helix.

returns to normal within three cross-sections, however the slightly more constrained A-Ser has a more diffuse tight twist region that extends up to 125°, but stays elevated and more variable for six cross-sections. This more constrained area around the mutation would make it more inflexible, leading to the observed lower melting point of 29.5 °C, compared to A-Ala at 34.0 °C.

4.2.3 Triple Mutant Helices

Unlike A-Ala and A-Ser, 3-Ala and 3-Ser fold into the designed A'A'B' composition and register triple helix. Both have four peaks of equal volume in the ^{1}H,^{15}N-

HSQC indicating four distinct isotopic labels in the trimer (labeled G1-G4) as expected in the A'A'B' triple helix and analogous to the base system. These systems, however, are significantly less stable and less ordered than the base triple helix, having melting points of 19 °C and 18 °C respectively. This is lower than either A-Ala or A-Ser, and have the lowest melting point of any of the peptides investigated (Table 4.1, Fig. 4.3).

3-Ala has relaxation values (5.3, 2.4, 3.0, and 4.5) below that of A-Ala (7.7, 8.9, 11.9, 7.8, and 6.0), and substantially lower than the base triple helix, indicating a very disordered region around the mutation. Additionally there is significantly less helical population compared to the monomer (0.19 vs. 0.81), lower even than A-Ala (Table 4.1). 3-Ala also appears to adopt a second minor register as evidenced by the presence of an additional set of triple helix peaks in the $^{1}H,^{15}N$-HSQC (Fig. 4.2). There are a number of possibilities for the alternate register, but because of low folded populations and overlapping chemical shifts, further characterization was not possible.

The model of the 3-Ala A'A'B' helix was surprisingly well-folded considering this helix must incorporate three mutations, with only one lost hydrogen bond, and nine formed charge pairs. The lost hydrogen bonds and charge pairs are in the mutated region, which suggests a local disruption that compensates for the mutations. The helical twist of A-Ala agrees with this hypothesis with a shift towards larger twist angles, up to 130°, around the mutated region extending for five cross-sections. This suggests a perturbed and inflexible helix that has difficulty compensating for the three mutations, leading to the low melting point.

3-Ser also retained composition and register control to form the A'A'B' triple helix and also showed a drop in relaxation values (8.3, 4.8, 5.3, and 5.8) when compared

to A-Ser indicating greater conformational freedom (Table 4.1). 3-Ser had a fraction of trimer to monomer of 0.17 which is similar to A-Ser. This is notable because 3-Ser has three mutated amino acids and still is able to fold a comparable amount of triple helices as A-Ser which only has one mutation. In the 3-Ser A'A'B' model, two hydrogen bonds and one charge pair is lost, similar to 3-Ala model. However, the serine OH side chains contribute an additional four hydrogen bonds, between the adjacent OH groups and with the peptide backbone leading to a more constrained mutated region. The helical twist data also suggests a more constrained mutated region with values extending over 130° and under 80°, with the disruption extending over five cross-sections. This suggests a highly perturbed yet inflexible structure, which is compatible with the low melting point.

4.3 Conclusions

The designed triple helical systems presented in this paper show that additional mutations in the constituent peptide strands of a triple helix result in a drop in melting temperature. In all cases the triple helix folding is affected by two major factors: 1) the number of mutations in the triple helix and 2) the stability and inflexibility conferred by axial charge pairs. The battle between these factors is clear in the A-Ala and A-Ser cases where the designed A'A'B triple helix would have ten axial charge pairs and two mutations, and nicely satisfy the 2:1 mixing ratio; the triple helix that actually folds, however, is the A'BB triple helix, with only five satisfied charge pairs but also only one mutation and still a significant portion of unfolded single chains. The number of mutations present clearly outweighs the stabilization from charge pairs and therefore the mutation directs the composition of the triple helix. In contrast, the 3-Ala and 3-Ser systems only have mutated peptides available which forces the formation of a triple helix

with three mutations, and the one with the most axial charge pairs, A'A'B', forms. There is no alternative system where fewer mutations could be incorporated. Despite retention of the designed composition and register, these are weakly folded systems with the largest amount of unfolded peptide. In the folded systems, the charge pairs are in direct opposition to the flexibility needed to incorporate the mutations. This tension leads to wild variation in the helical twist, and an overall inflexible and unstable region. However, when fewer formed charge pairs are present, such as the comparison between B-Ala and A-Ala with seven and five formed charge pairs respectively, some of the tension on the helix is lifted, allowing for a locally disrupted helix which can normalize over less cross-sections than the more constrained B-Ala helix.

4.4 Experimental

4.4.1 Peptide Synthesis

Peptide synthesis was performed on an Advanced Chemtech Apex 396 peptide synthesizer utilizing standard solid phase peptide synthesis techniques with FMOC protecting groups. The peptides were synthesized on a 0.15 mM scale on Rink amide MBHA resin (0.37 mmol/gram loading). Deprotection cycles were carried out with a 25% piperidine in dimethylformamide solution. Coupling cycles employed a 1:4:4:6 ratio of resin : amino acid : HATU : diisopropylethylamine (DiEA). For coupling ^{15}N-glycine isotopic labels, this ratio was reduced to 1:2:2:3. Acetylation at the N-terminus was performed twice with a mixture of resin : DiEA : acetic anhydride (1:10:5) in dichloromethane. A standard cleavage cocktail containing EDT :H_2O : TIPS (1:1:0.4) in 20 mL trifluoroacetic acid (TFA) was performed for three hours to remove the peptides from the resin. The TFA was removed using rotary evaporation, and then the product was

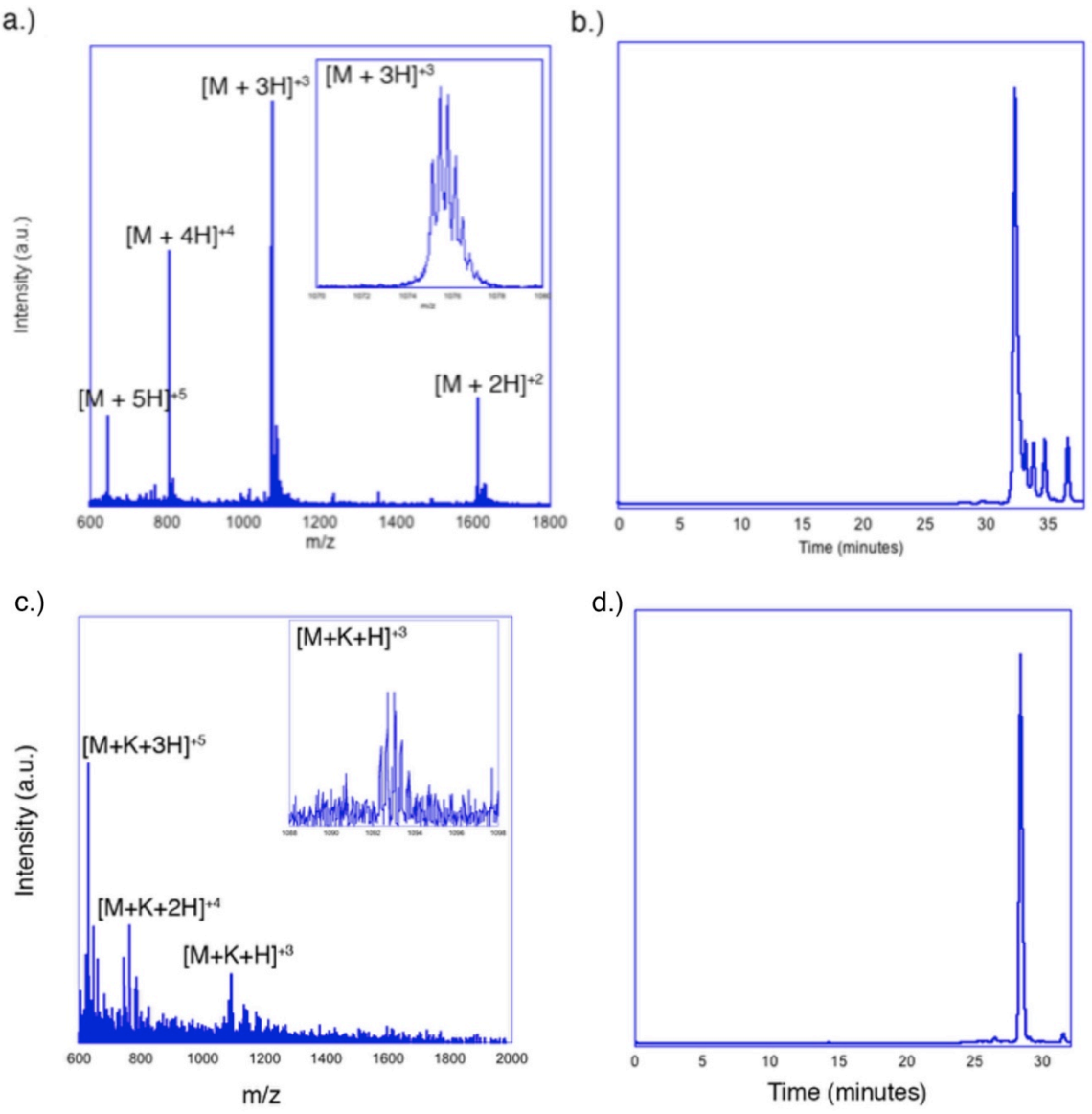


Figure 4.8. a) ESI-TOF MS of the purified fractions (Predicted MW $[M+H]^{+}$= 3324.49, Observed 3224.67) and b) HPLC trace of the A-Ala A peptide. c) ESI-TOF MS of the purified fractions (Predicted MW $[M+K+H]^{3+}$= 3279.50, Observed 3279.23) and d) HPLC trace of the A-Ser A peptide.

precipitated out with cold diethyl ether. The precipitate was pelleted using centrifugation and the ether wash repeated two additional times.

4.4.2 Peptide purification

The peptides were purified with a combination of dialysis and HPLC. Dialysis was performed using SpectraPor tubing with a 500-1000 MW cutoff for 3 days. Then the solution was diluted to roughly 5 mg of peptide/ 1 mL H_2O, and purified through reverse phase high pressure liquid chromatography (HPLC). HPLC was performed on a Varian PrepStar220 with a Phenomenex Jupiter 4μ Proteo 90 Å column. A gradient of 1% acetonitrile (with 0.05% TFA) in water (with 0.05% TFA) was used and the main peak was confirmed using electrospray ionization time-of-flight mass spectrometry (ESI TOF-MS). MS and HPLC spectra (Fig. 4.8). The TFA and acetonitrile were removed via rotary evaporation, and the sample was frozen and lyophilized to give pure peptide powder.

4.4.3 Sample preparation

Stock solutions of peptides were prepared with Milli-Q water and the pH was adjusted to 7 with sodium hydroxide. For circular dichroism the samples were diluted to 0.3 mM peptide with 10 mM phosphate, and for NMR the samples were prepared at 3 mM total peptide concentration and 10 mM phosphate. These solutions were annealed at 85 °C and allowed to incubate for 1 week at room temperature and 1 week at 4 °C.

4.4.4 Circular Dichroism Spectroscopy

A Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller was used to record all secondary structure spectra and thermal unfolding studies. Spectra were collected between 190 and 250 nm. Melting experiments monitored the maxima around 225 nm from 5 – 85 °C at a heating rate of 10 °C/hour, and the minima of the first derivative of the melting curve was used to determine the melting temperature of the sample. The molar residual ellipticity (MRE), shown as [θ], was calculated from the recorded ellipticity using:

$$[\theta] = \frac{\theta}{c \times l \times nr \times 10}$$

In the equation above θ represents the ellipticity, m is the molecular weight (g/mol), c is the sample concentration (M), l is the pathlength (cm) and n_r is the number of residues in the peptide. The first derivatives of the melting curves are available in Fig. 4.3, and the melting points are given in Table 4.1.

4.4.5 NMR Spectroscopy

All TOCSY, and ^{1}H,^{15}N-HSQC and ^{1}H,^{15}N-NOESY-HSQC experiments were performed on a Varian spectrometer of 800 MHz containing a cryogenic probe. The nitrogen carrier frequency was set at 110 ppm, while the proton carrier frequency set to the water signal. All experiments were performed at 10 °C. Chemical shift assignments of residues were made using TOCSY spectra. For TOSCY spectra 1120 x 512 complex points and 128 scans were acquired; the mixing time was 100 ms and sweep width 12 kHz in both dimensions. The ^{1}H,^{15}N-HSQC were collected with 1040 x 256 complex points for a total of 16 scans. A sweep width of 10 kHz in the proton and 2432 Hz in the nitrogen dimension was used. To compare relative peak volumes of folded and unfolded states, for each ^{1}H,^{15}N-HSQC of an AAB heterotrimer the total peak volume was set to 1. Because of the difference in the relaxation rate of the monomer and folded helix no quantitative relationship between the populations can be determined from this ratio, but a general comparison between the mutation types may be made (Table 1). A 2D version of the 3D ^{1}H,^{15}N-NOESY-HSQC collected with 100 ms mixing time, 64 scans, 1200 x 256 complex points; the spectral window was 12 kHz in the direct dimension and 12 kHz for the indirect dimension. ^{15}N relaxation experiments were performed on 3 mM samples as previously described on a 500 MHz Varian spectrometer.[13] Data was processed and

analyzed using NMRdraw,[15] and ccpnmr[16] processing suites. Additional spectra including ^{1}H,^{15}N-NOESY-HSQCs and TOCSY, along with a discussion of the assignment process are available in Appendix 2.

4.4.6 Molecular Modeling

Models were created by first preparing a basic structure with the appropriate sequence, and expected composition and register. For the B-Ala, B-Ser, 3-Ala and 3-Ser helices an AAB composition and register was used. For A-Ala and A-Ser an ABB composition and register was used. Constraints were created for the heterotrimers from their assigned ^{1}H,^{15}N-NOESY-HSQC spectrum. These constraints were used to model the heterotrimers using the Rosetta Relax application.[17, 18] Negative constraints were created for situations where constraints were present in the base structure, but not the mutated structure. Negative constraints were not used for the three chain mutants because their initial structure deviated significantly from the base structure. Positive correlations were set for 0-4 Å and negative from 6-10 Å. Once a set of structures was created, they were analyzed for compliance to the constraints. Positive correlations were considered satisfied if they were below 5 Å and negative correlations were considered satisfied if they were above 5 Å. The most compliant structure was then iterated until all constraints were satisfied. If any constraints were still unsatisfied after multiple iterations the standard deviation of that constraint was decreased to increase the penalty for longer distances. The final structures presented here are compliant to all NMR constraints. The constraints used are available in Appendix 1.

4.4.7 Helical Twist

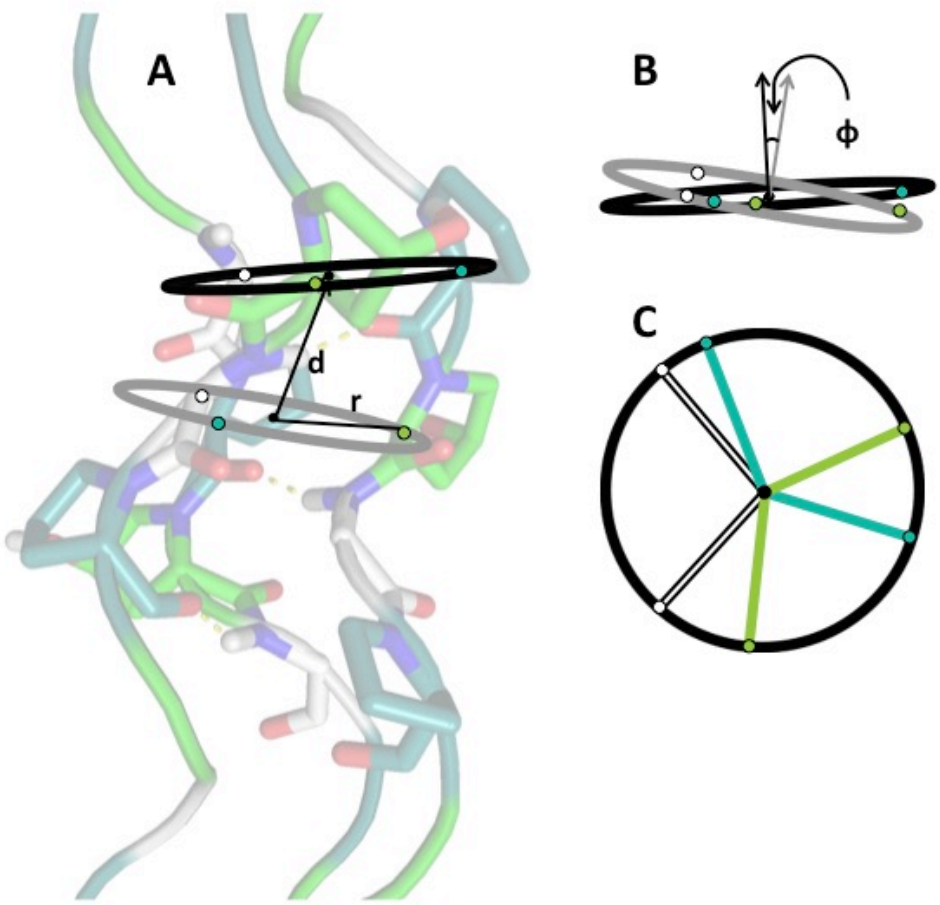


Figure 4.9. Panel A illustrate two adjacent triads composed of the three α carbons in each cross section. The displacement between the centroids is shown as d and the radius as r. Panel B shows the two triad centroids overlaid to illustrate the tilt angle between them. In panel C, the triads are viewed from above and have been tilted such that they lie in the same plane. Illustrated are the angles of interest in each respective color green, turquoise, and white.

A program was written in Java with a J/Link to Mathematica[19] to perform the twist measure calculations in this study. For each residue the coordinates of the alpha carbon were considered for the purpose of conducting calculations. The program was used to extract alpha carbon coordinates from two adjacent triads on the helical axis and perform a series of transformation calculations on these coordinates. The calculations run to assess the measure of twist along the helical axis were done by taking cross-sections of

the helix, perpendicular to the helical axis, containing an amino acid from each of the three strands of the triple helix (illustrated in Fig. 4.9). This ensures that each of the amino acid residues in a cross section will occupy a different position in the conventional Xaa-Yaa-Gly structure of a collagen triple helix due to the staggered nature of helical collagen. From here, a cross section containing Xaa, Yaa, and Gly position amino acids will be referred to as a triad to differentiate from a conventional Xaa-Yaa-Gly *triplet*. The coordinates of the alpha carbons of each residue were used to construct the coordinate matrices for each triad.

The Kabsch twist values [21] were obtained in a similar manner to a recent report [22] in which he utilized the CNS package which uses the Kabsch least-squares fitting algorithm to calculate a least-squares superposition matrix and translation vector from which he extracted helical twist and translation. In our study however, we utilized a script in Java to call a Mathematica notebook to calculate the covariance matrix of two triads adjacent on the helical axis. Mathematica was then used to perform a Singular Value Decomposition (SVD) on the covariance matrix and the result was used to calculate the optimal rotation matrix (**U**) according to the Kabsch algorithm. From this the translation vector $\vec{\boldsymbol{u}}$ can be extracted as follows:

$$for: \quad \boldsymbol{U} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$\vec{\boldsymbol{u}} = \pm \begin{bmatrix} h - f \\ c - g \\ d - b \end{bmatrix}$$

and the angle of rotation θ is embedded in the rotation matrix **U** as follows:

$$for: \quad \vec{\boldsymbol{u}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\boldsymbol{U} = \begin{bmatrix} x^2(1-cos\,\theta)+cos\,\theta & xy(1-cos\,\theta)-z\,sin\,\theta & xz(1-\cos\theta)+y\sin\theta \\ xy(1-\cos\theta)+z\sin\theta & y^2(1-cos\,\theta)+cos\,\theta & yz(1-\cos\theta)-x\sin\theta \\ xz(1-\cos\theta)-y\sin\theta & yz(1-\cos\theta)+x\sin\square & z^2(1-cos\,\theta)+cos\,\theta \end{bmatrix}$$

or, more simply:

$$Trace(\boldsymbol{U}) = 1 + 2\cos\theta$$

Applying this optimal rotation matrix and translation vector to the points in one triad will not exactly overlay with the points of the adjacent triad due to variations in the collagen helix due to variables like the identity of the adjacent triads, the position of each triad in the triple helix, etc.

The radius and displacement values were calculated with respect to the centroid of each triad. The centroid was calculated by taking the average of all of the x, y, and z coordinates in a triad. The displacement values are the distance from each residue to the centroid and the radius is the average of these three values.

The translation between triads was calculated as the distance between the centroids of adjacent triads. The accompanying unit vector is the unit vector the points from the first centroid to the next and was calculated by taking the difference of each set of coordinates and dividing by the magnitude of the translation.

The angle of tilt between two adjacent triads was calculated by first shifting the centroid of each triad to the origin by subtracting the coordinates of the centroid from the

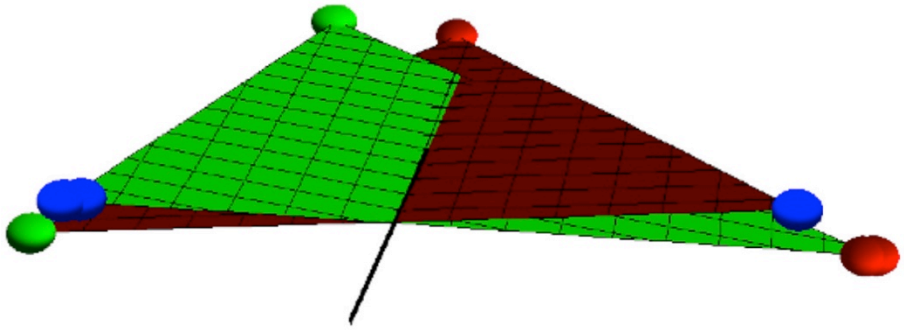


Figure 4.10. Illustration of the intersection of planes created by each triad.

coordinates of each individual alpha carbon. This ensured that the centroids of both triads were coincident with one-another and that there exists a single line at the intersection of the planes created by each triad as shown in Figure 4.10. The unit vector pointing in the direction of this line was calculated by taking the cross product of the vectors that point perpendicular from each triad's plane. These vectors were obtained by using Mathematica to solve for the equation of the plane through the alpha carbons of each triad and using the x, y, and z coefficients as the coordinates for these vectors. The tilting angle between the planes was calculated according to the following:

$$for\ perpendicular\ vectors\ \vec{r}\ and\ \vec{s}$$

$$\cos\theta = \frac{\vec{r}\cdot\vec{s}}{\|\vec{r}\|*\|\vec{s}\|} \qquad eq.\ 5$$

For the twisting angle to be calculated, first the tilting transformation had to be done to the second triad. This was done by first formulating a rotation matrix of the same form as equation 3 by using the unit vector of the line of intersection and the angle calculated in equation 5 and then left multiplying each set of coordinates of the second triad by this rotation matrix. The angle of twist was then obtained by measuring the angle

between the coordinates of each residue on the first triad, the centroid, and the coordinates of the corresponding residue on the second triad.

4.5 References

(1) Ramachandran, G. N.; Sasisekharan V. *Nature* **1961**, *190*, 1004-1005

(2) Rich, A.; Crick, F. H. C. *J. Mol. Bio.* **1961**, *3*, 483–506

(3) Kielty, C. M.; Hopkinson, I. In *Connective Tissue and Its Heritable Disorders*, 2nd ed. Royce, P. M., Steinmann, B., Eds.; John Wiley & Sons Inc.: Hoboken, NJ, 2003; p 159.

(4) Marini, J. C.; Forlino, A.; Cabral, W. A.; Barnes, A. M.; San Antonio, J. D.; Milgrom, S.; Hyland, J. C.; Korkko, J.; Prockop, D. J.; De Paepe, A.; Coucke, P.; Symoens, S.; Glorieux, F. H.; Roughley, P. J.; Lund, A. M.; Kuurila-Svahn, K.; Hartikka, H.; Cohn, D. H.; Krakow, D.; Mottes, M.; Schwarze, U.; Chen, D.; Yang, K.; Kuslich, C.; Troendle, J.; Dalgleish, R.; and Byers, P. H. *Hum. Mutat.* **2007**, *28*, 209–221

(5) Byers, P.H.; Wallis G. A.; Willing, M. C. *J. Med. Genet.* **1991**, *28*, 433-442

(6) Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. *Proc. Natl. Acad. Sci.* **2000**, *97*, 4273–4278

(7) Radmer, R. J.; Klein, T. E. *Biochemistry* **2004**, *43*, 5314–5323

(8) Brodsky, B.; Thiagarajan, G.; Madhan, B.; Kar, K. *Biopolymers* **2008**, *89*, 345–353

(9) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75–81

(10) Bella, J.; Brodsky, B.; Berman, H. M. *Structure **1995**, 3*, 893–90

(11) Li, Y.; Brodsky, B.; Baum, J. *J. Biol. Chem.* **2009**, *284*, 20660–20667

(12) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7509–7515

(13) Acevedo-Jake, A. M.; Clements, K. A.; Hartgerink, J. D. *Biomacromolecules* **2016** *17*, 914–921

(14) Jalan, A.A.; Hartgerink, J.D. *Biomacromolecules*. **2013**, *14*, 179-185

(15) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293

(16) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* **2005**, *59*, 687–696

(17) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. *Meth. Enzymol.* **2004**, *383*, 66–93

(18) Leaver-Fay, A.; O'Meara, M. J.; Tyka, M.; Jacak; R., Song, Y.; Kellogg, E. H.; Thompson, J.; Davis, I. W.; Pache, R. A.; Lyskov, S.; Gray, J. J.; Kortemme, T.; Richardson, J. S.; Havranek, J. J.; Snoeyink, J.; Baker, D.; Kuhlman, B. *Meth. Enzymol.* **2013**, *523*, 109–143

(19) Wolfram Research, Inc. *Mathematica* **2016**, 11 Ed., Champaign, IL

(20) Kabsch, W. *Acta Crystallogr.* **1976**, *32*, 922–923

(21) Bella, J. *J. Struct. Biol.* **2010**, *170*, 377–391

Chapter 5: Conclusions

Collagen is a ubiquitous protein that functions as a scaffold, signaling protein, and adhesion point for cells. Much is known about collagen's functions, but less is known about the supramolecular forces affecting the triple helix. I have presented research that elucidates the relationship of charge pairs on stability and have given insight to the impact of mutations on composition and register control in the triple helix.

In a canonical collagen triple helix, three peptides self-assemble into a supercoiled motif with a one amino acid offset between the peptide chains. Design of triple helices that contain more than one residue offset is lucrative, as it leaves the non-covalent interactions unsatisfied at the termini and renders the termini "sticky" to further self-assemble into collagen-like nanofibers. In the second chapter I, with coworkers, used lysine−glutamate axial salt-bridges to design a heterotrimeric collagen triple helix, ABC-1, containing a non-canonical offset of four residues between the peptide chains. A second heterotrimer, ABC-2, also stabilized by axial salt-bridges, was designed containing a canonical one-amino-acid offset to facilitate comparison of structure and stability by CD and NMR. ABC-1 and ABC- 2 demonstrate our ability to modulate chain offset in a collagen triple helix. These "stickier" structures require a very delicate balance of stabilizing charge pairs in the offset helix, and destabilizing interactions in the blunt helix because both species will be possible in solution. This lays the groundwork to design longer, and therefore stickier, offsets allowing access to a new class of collagen-related nanostructures.

In Chapters 3 and 4 I discussed Osteogenesis imperfecta (OI) which is a disease caused primarily by mutations of glycine in the standard $(Xaa\text{-}Yaa\text{-}Gly)_n$ repeat of a type

I collagen triple helix. Type I collagen is an AAB heterotrimer, which means that, depending on whether the A or B chain is mutated, the glycine substitution will appear once or twice. In Chapter 3, designed axial charged pairs were used to self-assemble an AAB triple helix with controlled composition and register. Then a single glycine of the B chain with alanine, serine, valine, aspartate, or arginine was substituted and assessed for the impact on the structure and folding of this OI mimic by CD, NMR, and restraint-guided modeling. It was found that alanine and serine substitutions are tolerated, resulting in localized disruptions to the triple helix structure, while bulkier amino acids result in alternatively folded structures. Models guided by NMR of the alanine and serine substitutions show composition and register control with the loss of 5 and 2, respectively, hydrogen bonds around the mutation site, and wide variation in the helical twist due to the mutation and the conformational demands of the charge pairs. Larger amino acids, such as valine, aspartate, and arginine, result in loss of composition and register control, where the homotrimers of the A peptide folds instead of the designed AAB helix. This work demonstrates the potential of axial charged pairs to control the structure of low stability triple helices and also helps to elucidate the structure and folding challenges associated with OI-type mutations in collagen.

In Chapter 4 I presented a series of triple helices with mutations to a composition- and register-controlled AAB helix where one of the requisite glycines in the A chain of the triple helix is changed to serine or alanine. There is a loss of compositional control when the A chain is mutated, resulting in an A′BB composition that minimizes the number of mutations included in the triple helix. This demonstrates that the stabilization generated by charge pairs is insufficient to compensate for the destabilization of the

double mutation. A combination of NMR and molecular modeling revealed that only one hydrogen bond was lost in each of the alanine and serine cases and all possible charge pairs were formed. However, when both A and B chains are mutated and no non-mutated peptide chains are available, the designed A′A′B′ composition is reestablished and the resulting triple helix is surprisingly well folded while compensating for three mutations. This work shows the ability of the mutations to influence and alter the composition and register of the collagen triple helix.

Globally, the OI studies are still at a low level of relevance to biological systems. In order to increase the relevance of the model systems they need to be able to integrate sequences where the mutations occur in natural collagen. However, the natural sequences disrupt the composition and register control, which can be exacerbated by the mutations. In the beginning of this project five alternate sequences of the model were attempted, each containing natural sequences, however these models were either unable to fold, or would fold in the incorrect composition and register with all mutations. While our model triple helical systems give information about the composition and register created, that control is artificial, and the twist created is a combination of both the mutation and charge pairs.

Future studies may be able to overcome these difficulties when better understanding of pairwise interactions in collagen is gained allowing for more effective design of the host peptide sequences. Alternatively, bacterial expression of a natural sequence with the specific mutation could be used study the specific effect of that mutation, but heterotrimeric, bacterially expressed peptides with natural sequences have not been shown to exhibit composition control. While the designed triple helices used in

my study are a step in the direction of biological relevance much improvement is needed until biological and clinicial significance, not just supramolecular folding principles, can be ascertained from the results.

Appendix 1: Modeling Constraints

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
45B	2HA	45B	positive	1
45B	1HA	45B	positive	1
45B	1HB	72C	positive	1
45B	1HD	14A	positive	1
75C	HA	74C	positive	1
75C	H	76C	positive	1
75C	1HA	75C	positive	1
75C	1HA	18A	positive	1
75C	1HD	19A	positive	1
75C	1HD	44B	positive	1
75C	1HA	45B	positive	1
66C	HA	65C	positive	1
66C	2HA	66C	positive	1
66C	1HA	66C	positive	1
66C	1HD	67C	positive	1
66C	1HD	37B	positive	1
66C	2HB	65C	positive	1
66C	1HB	37B	positive	1
66C	1HB	65C	positive	1
66C	1HD	38B	positive	1
15A	1HA	15A	positive	1
15A	2HA	45B	positive	1
15A	1HD	43B	positive	1
15A	1HB	72C	negative	0.75
15A	2HB	72C	negative	0.75
15A	3HB	72C	negative	0.75
45B	HA	72C	negative	0.75
45B	HA	74C	negative	0.75
15A	HA	44B	negative	0.75
15A	HA	16A	negative	0.75
15A	HA	71C	negative	0.75

Table A1.1. Constraints used for B-ala modeling

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
75C	1HB	74C	positive	1
75C	1HA	75C	positive	1
75C	1HB	76C	positive	1
75C	2HB	76C	positive	1
75C	1HD	19A	positive	1
15A	2HB	72C	positive	0.5
15A	1HD	14A	positive	1
15A	1HA	15A	positive	1
66C	H	8A	positive	1
66C	HA	8A	positive	1
66C	1HB	8A	positive	1
66C	1HD	8A	positive	1
66C	1HB	65C	positive	1
66C	1HA	66C	positive	1
66C	2HA	66C	positive	1
45B	H	72C	positive	0.5
45B	HG	72C	positive	1
45B	1HD	73C	positive	1
45B	HA	74C	positive	1
45B	1HB	74C	positive	0.5
45B	H	75C	positive	1
45B	1HA	75C	positive	0.25
45B	1HA	15A	positive	1
45B	H	44B	positive	1
45B	HA	44B	positive	1
45B	1HA	45B	positive	1
45B	2HA	45B	positive	1
66C	HA	37B	negative	1
66C	HA	67C	negative	1
66C	HA	7A	negative	1
75C	HA	73C	negative	1
75C	HA	77C	negative	1
75C	HA	19A	negative	1
15A	HA	71C	negative	1
15A	HA	73C	negative	1
15A	HA	74C	negative	1
45B	HA	47B	negative	1

Table A1.2. Constraints used for B-ser modeling.

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
45B	HA	44B	positive	1
45B	2HA	45B	positive	1
45B	1HA	45B	positive	1
45B	1HB	44B	positive	1
45B	1HD	14A	positive	1
75C	HA	74C	positive	1
75C	1HA	75C	positive	1
75C	HA	16A	positive	1
75C	1HA	45B	positive	1
75C	1HB	74C	positive	1
75C	1HB	44B	positive	1
75C	H	46B	positive	1
75C	H	18A	positive	1
66C	HA	65C	positive	1
66C	HA	8A	positive	1
66C	2HA	66C	positive	1
66C	1HA	66C	positive	1
66C	2HD	67C	positive	1
66C	1HD	7A	positive	1
66C	1HD	67C	positive	1
66C	2HB	65C	positive	1
66C	1HB	65C	positive	1
66C	H	8A	positive	1
66C	H	36B	positive	1
15A	HA	44B	positive	1
15A	1HA	72C	positive	1
15A	2HA	45B	positive	1
15A	1HD	43B	positive	1
15A	1HB	73C	positive	1
15A	2HB	44B	positive	1
15A	1HB	44B	positive	1
66C	1HA	9A	negative	1
66C	HA	35B	negative	1
15A	HA	17A	negative	1
15A	HA	74C	negative	1
45B	1HA	18A	negative	1
75C	1HA	15A	negative	1
75C	HA	47B	negative	1
75C	HA	76C	negative	1

Table A1.3. Constraints used for the Base modeling

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
15A	1HD	14A	positive	1
15A	1HA	45B	positive	1
15A	1HA	15A	positive	1
15A	HA	14A	positive	1
15A	H	16A	positive	1
45B	H	46B	positive	1
45B	H	15A	positive	1
45B	HA	44B	positive	1
45B	1HA	15A	positive	1
45B	1HA	45B	positive	1
45B	2HA	45B	positive	1
45B	1HD	73C	positive	1
45B	1HB	44B	positive	1
45B	2HB	44B	positive	1
75C	1HB	74C	positive	1
75C	1HA	75C	positive	1
75C	2HA	75C	positive	1
75C	H	46B	positive	1
36B	1HB	35B	positive	1
36B	2HB	35B	positive	1
36B	1HD	7A	positive	1
36B	1HD	37B	positive	1
36B	1HA	36B	positive	1
36B	2HA	36B	positive	1
36B	HA	35B	positive	1
66C	1HB	65C	positive	1
66C	2HB	65C	positive	1
66C	1HB	7A	positive	1
66C	1HD	7A	positive	1
66C	1HD	67C	positive	1
66C	1HA	66C	positive	1
66C	2HA	66C	positive	1
15A	HA	13A	negative	0.5
45B	HA	16A	negative	0.5
75C	HA	73C	negative	0.5
66C	HA	64C	negative	0.5
36B	HA	64C	negative	0.5

Table A1.4. Constraints used for A-Ala modeling

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
75C	HA	74C	positive	1
75C	2HB	74C	positive	1
75C	1HB	74C	positive	1
75C	2HA	75C	positive	1
75C	1HA	75C	positive	1
75C	1HB	16A	positive	1
75C	2HA	45B	positive	1
75C	1HB	46B	positive	0.5
15A	HA	14A	positive	1
15A	1HA	15A	positive	1
15A	1HD	43B	positive	1
15A	1HA	45B	positive	1
36B	1HB	7A	positive	1
36B	HA	8A	positive	1
36B	HA	35B	positive	1
36B	1HB	35B	positive	1
36B	2HB	35B	positive	1
36B	1HA	36B	positive	1
66C	1HD	7A	positive	1
66C	2HB	7A	positive	1
66C	H	8A	positive	1
66C	HA	65C	positive	1
66C	2HB	65C	positive	1
66C	1HB	65C	positive	1
66C	HA	35B	positive	1
66C	2HA	36B	positive	1
66C	1HD	37B	positive	1
66C	1HA	66C	positive	1
66C	2HA	66C	positive	1
45B	1HD	73C	positive	1
45B	1HB	73C	positive	1
45B	1HD	14A	positive	0.5
45B	1HG	14A	positive	1
45B	1HA	15A	positive	1
45B	H	16A	positive	1
45B	HA	44B	positive	1
45B	1HB	44B	positive	1
45B	2HB	44B	positive	1
45B	1HA	45B	positive	1
45B	2HA	45B	positive	1
45B	H	46B	positive	1
66C	HA	67C	negative	1
66C	1HA	6A	negative	1
36B	HA	5A	negative	1
75C	1HA	15A	negative	0.5
75C	1HA	48B	negative	1
45B	HA	47B	negative	1
15A	HA	71C	negative	1
15A	HA	73C	negative	1

Table A1.5. Constraints used for A-Ser modeling.

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
45B	1HA	45B	positive	1
45B	2HA	45B	positive	1
45B	HA	44B	positive	1
45B	1HD	73C	positive	1
45B	1HG	14A	positive	1
45B	2HG	14A	positive	1
45B	1HB	44B	positive	1
45B	1HD	14A	positive	1
15A	1HA	45B	positive	1
15A	1HA	15A	positive	1
15A	2HA	15A	positive	1
15A	1HB	42B	positive	1
15A	H	42B	positive	0.5
15A	H	44B	positive	1
15A	1HB	43B	positive	1
15A	1HD	44B	positive	1
15A	HA	14A	positive	1
15A	1HD	43B	positive	1
75C	1HA	45B	positive	1
75C	1HA	75C	positive	1
75C	2HA	75C	positive	1
75C	H	76C	positive	1
75C	HA	74C	positive	1
75C	1HB	74C	positive	1
75C	1HD	19A	positive	1
75C	2HB	74C	positive	1
66C	1HA	66C	positive	1
66C	2HA	66C	positive	1
66C	HA	65C	positive	1
66C	1HB	65C	positive	1
66C	1HA	9A	positive	1

Table A1.6. Constraints used for 3-Ala modeling.

Isotopic Label	Partner Atom	Partner Residue	Type	Standard Deviation
15A	H	16A	positive	1
15A	HA	14A	positive	1
15A	1HA	15A	positive	1
15A	1HB	72C	positive	1
15A	1HD	14A	positive	1
15A	1HG	14A	positive	1
75C	1HA	75C	positive	1
75C	HA	74C	positive	1
75C	H	76C	positive	1
75C	1HB	74C	positive	1
75C	2HA	75C	positive	1
66C	1HA	66C	positive	1
66C	HA	65C	positive	1
66C	1HB	65C	positive	1
66C	2HB	65C	positive	1
66C	H	8A	positive	1
66C	1HD	67C	positive	1
66C	2HD	67C	positive	1
66C	1HB	37B	positive	1
45B	1HA	45B	positive	1
45B	2HA	45B	positive	1
45B	HA	44B	positive	1
45B	1HA	75C	positive	1
45B	1HB	16A	positive	1
45B	1HD	44B	positive	1

Table A1.7. Constraints used for 3-Ser modeling

Appendix 2: Supporting NMR Spectra

A2.1 Chapter 2

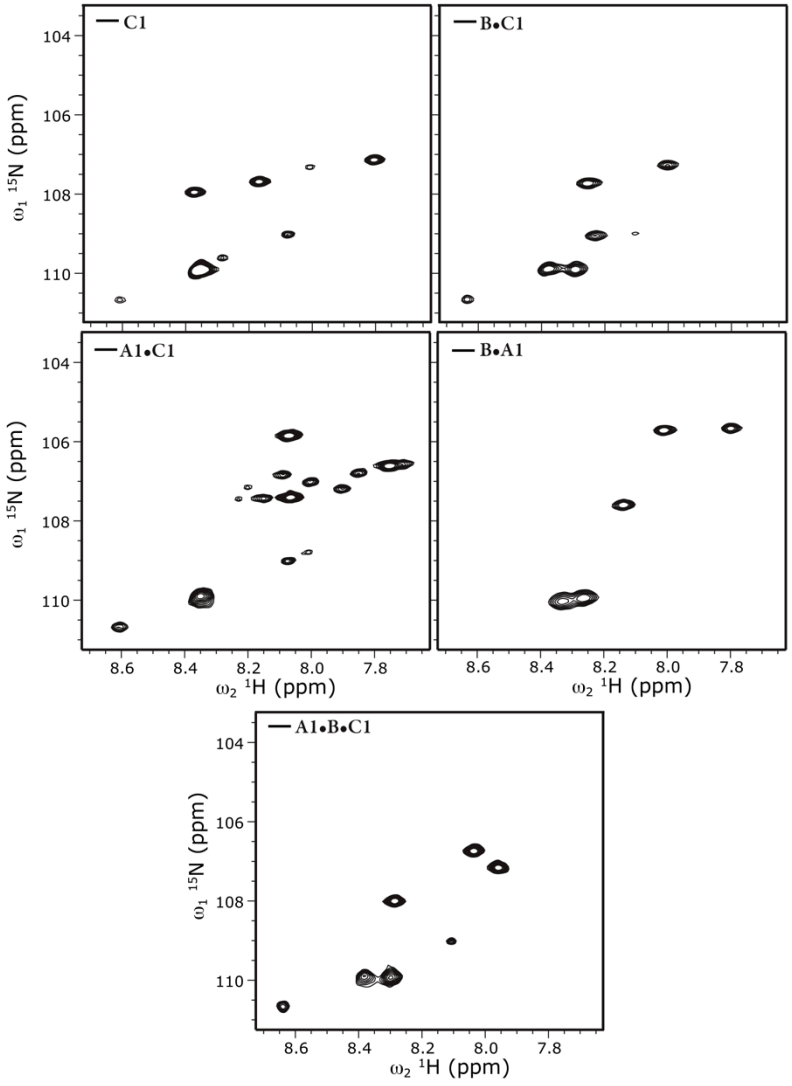


Figure A2.1. $^{1}H,^{15}N$-HSQC spectra of individual, binary and ternary peptide mixtures used for testing combinations ofA1,BandC1.

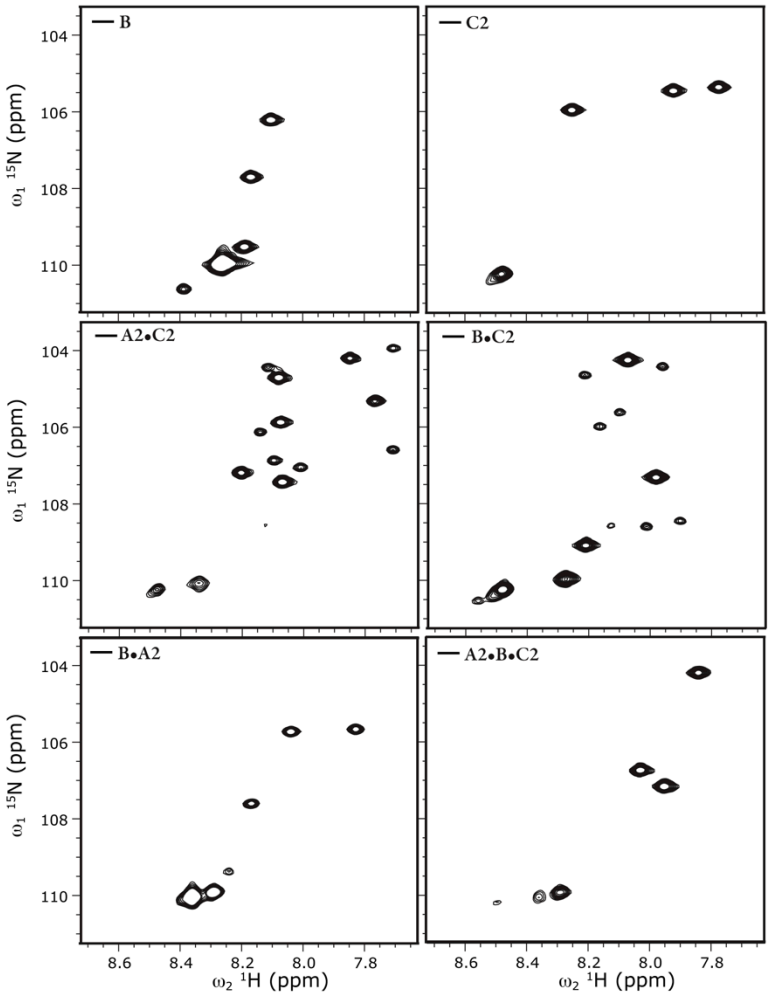


Figure A2.2. $^{1}H,^{15}N$-HSQC spectra of individual, binary and ternary peptide mixtures used for testing lateral combinations of A2, B and C2.

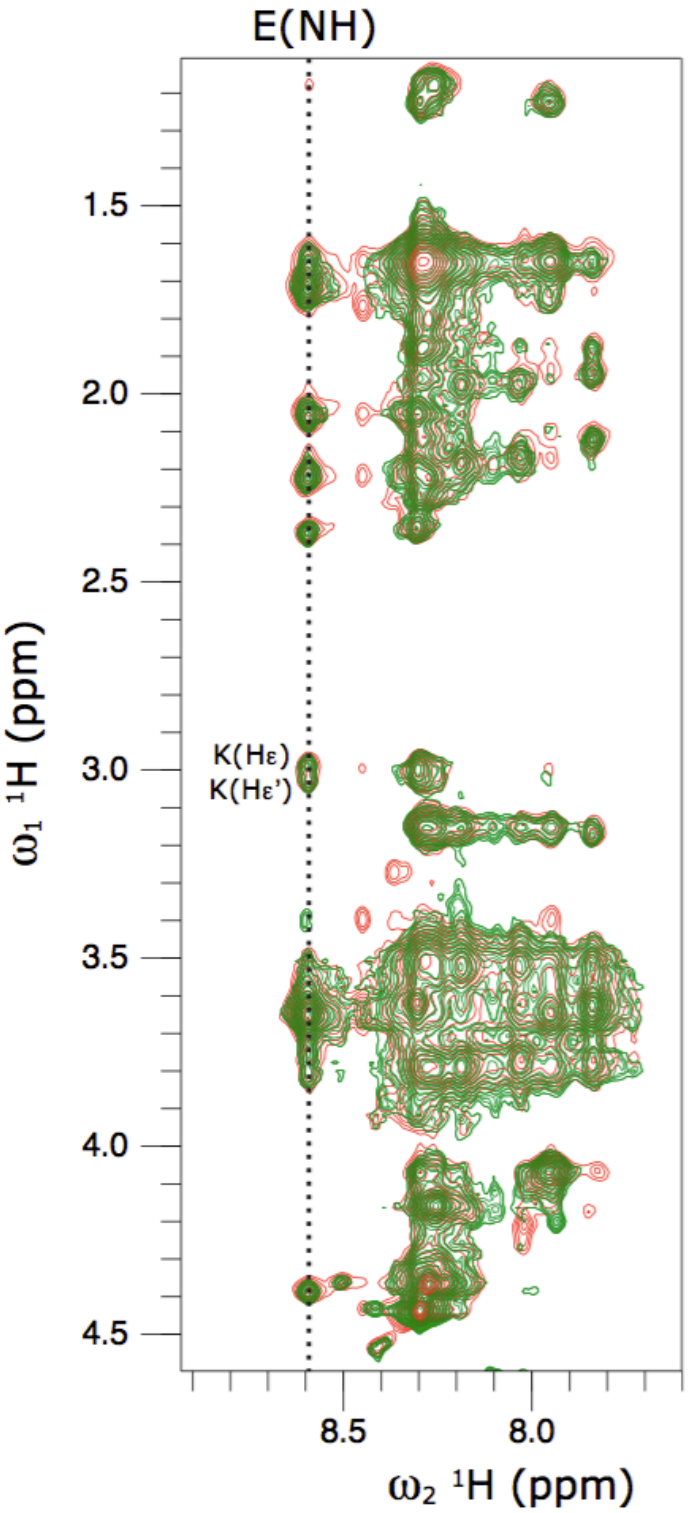


Figure A2.3. Amide region of the ^{1}H,^{1}H-NOESY spectrum of ABC-1 (green) and ABC-2 (red) indicating the similarity in the general cross-peak pattern and chemical shift. The *interpeptide* E(NH)-K(Hε) and E(NH)- K(Hε') cross peaks are also depicted.

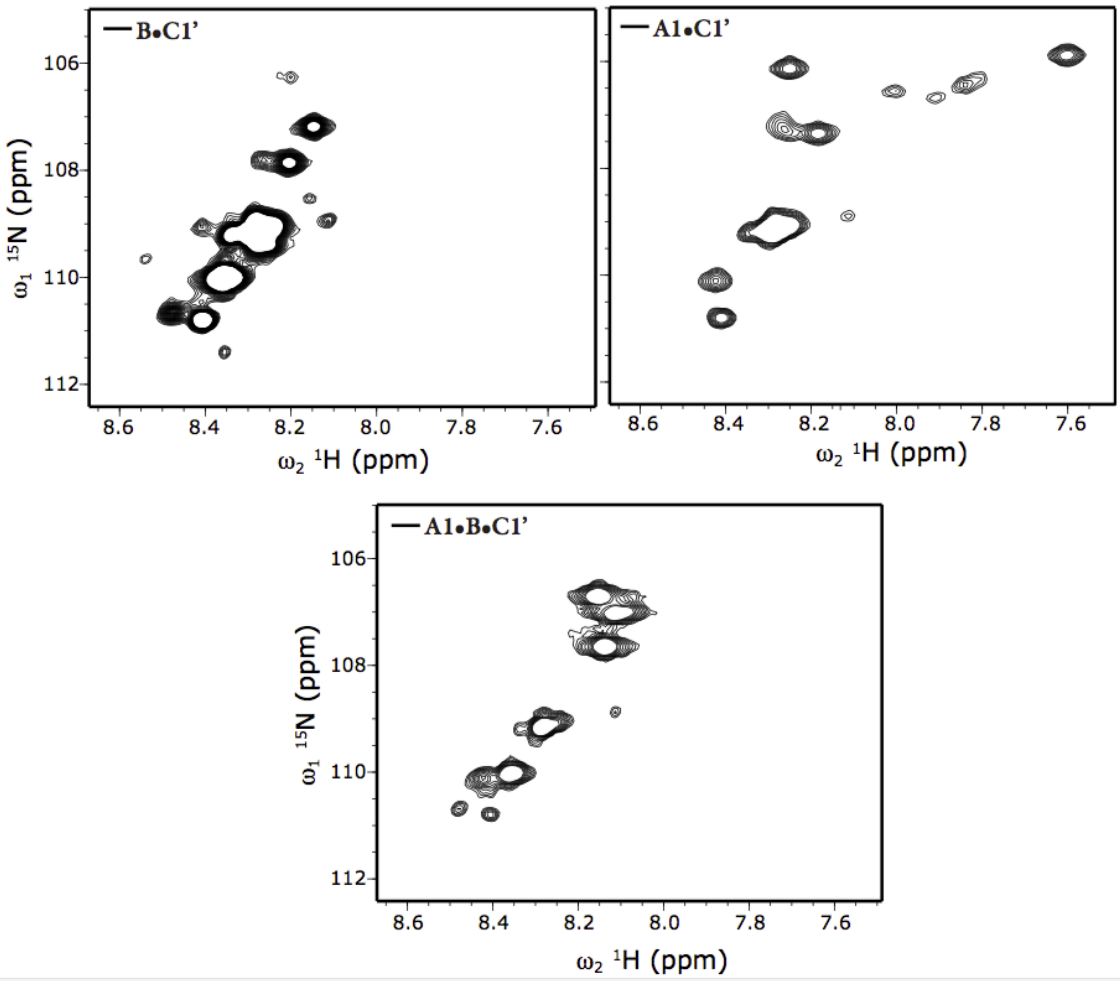


Figure A2.4. $^{1}H,^{15}N$-HSQC spectra of individual, binary and ternary peptide mixtures used for testing A1, B and C1'.

A2.2 Chapter 3

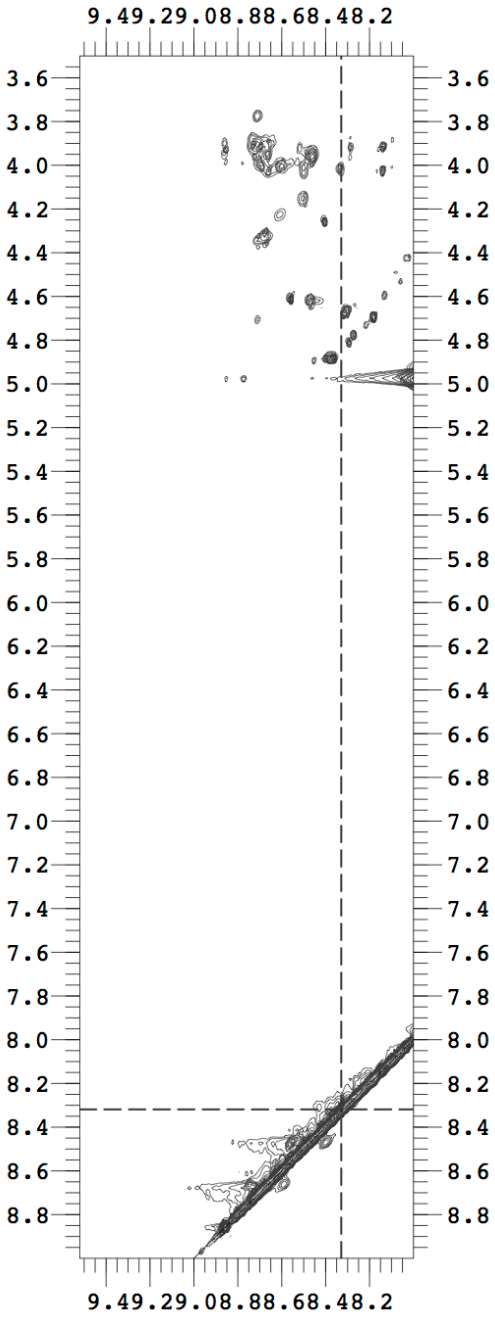


Figure A2.5. TOCSY figure showing connectivity for Gly3 in the Base system

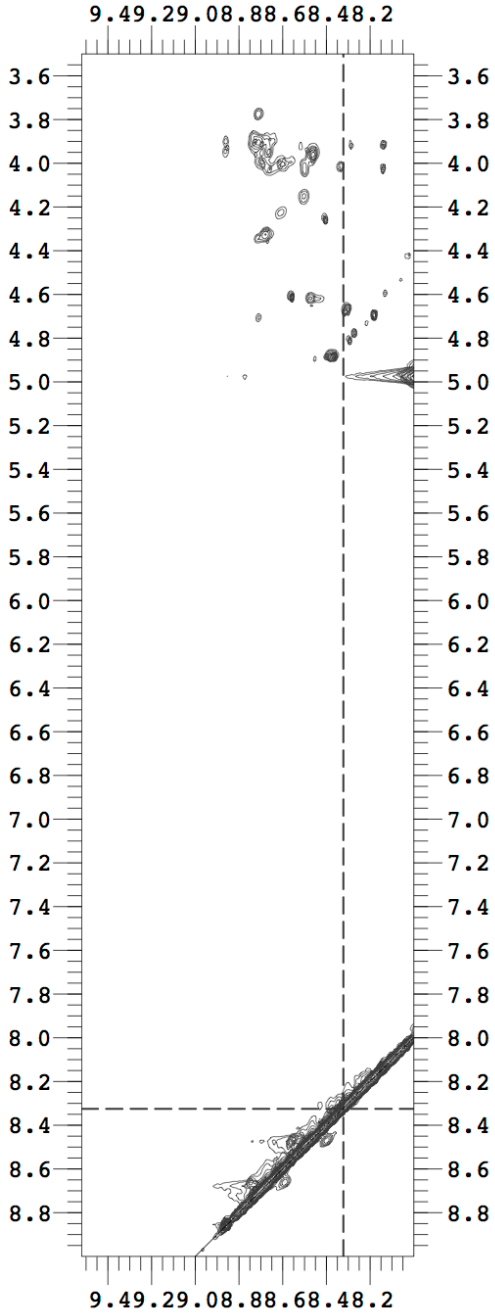


Figure A2.6. TOCSY figure showing connectivity for Gly3 in the B-Ala system.

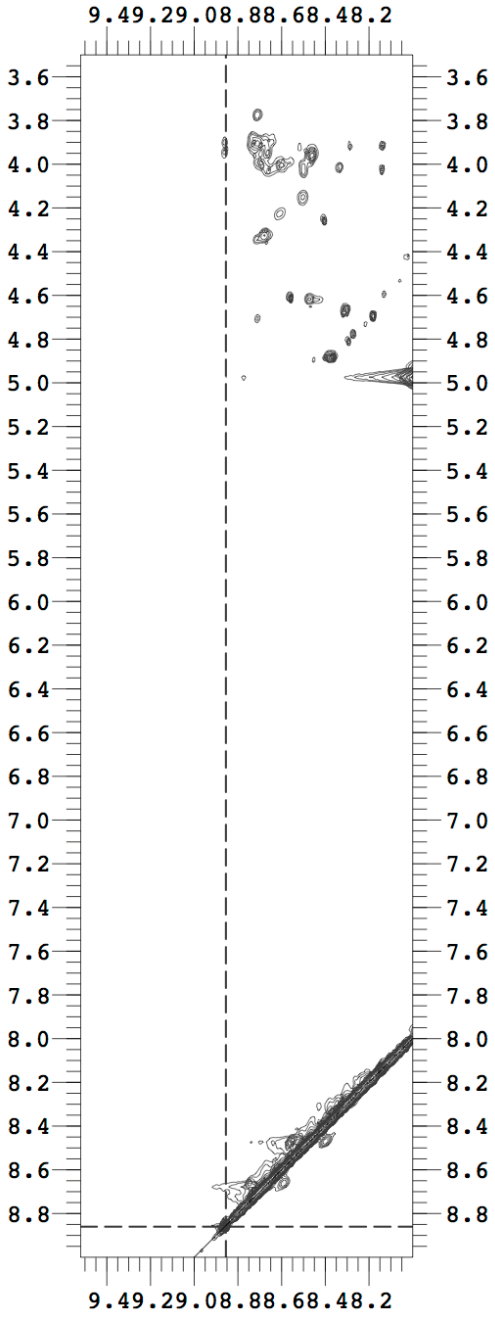


Figure A2.7. TOCSY figure showing connectivity for Gly3 in the B-Ser system.

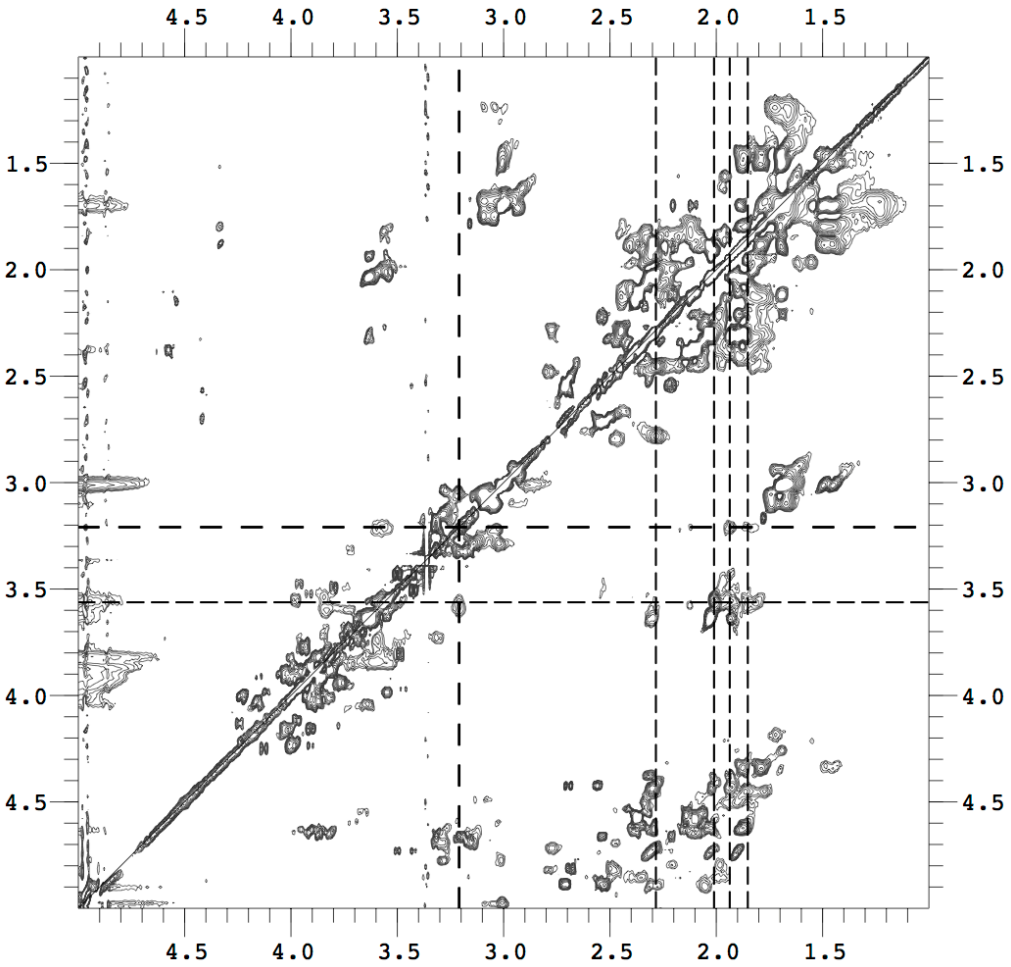


Figure A2.8. TOCSY figure showing connectivity for folded proline residues.

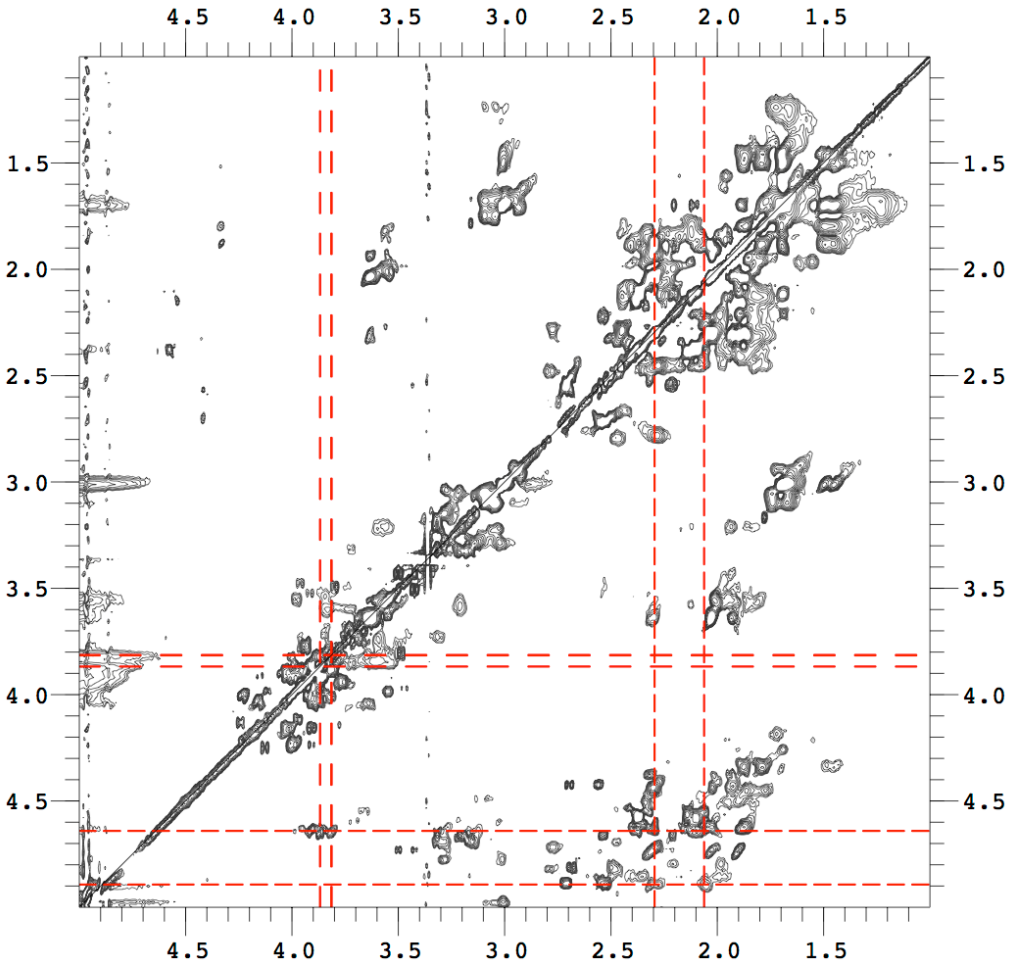


Figure A2.9. TOCSY figure showing connectivity for folded hydroxyproline residues.

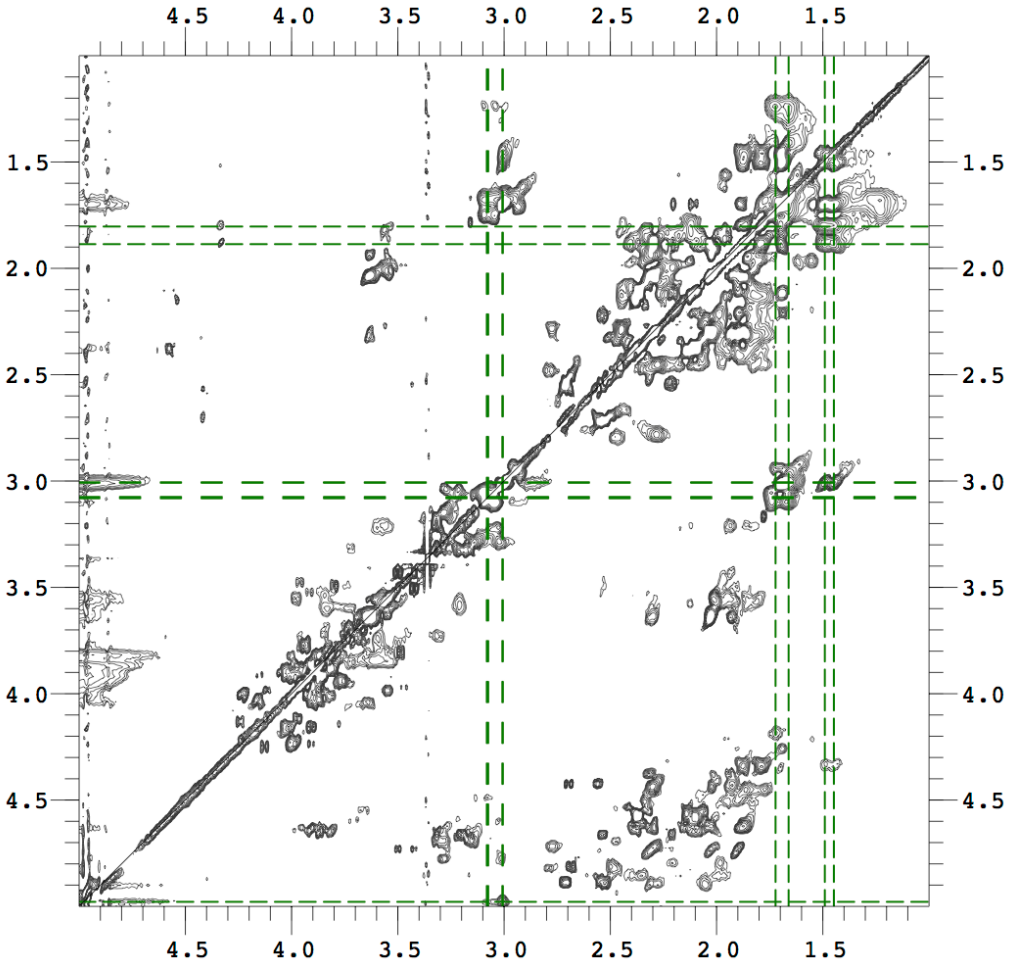


Figure A2.10. TOCSY figure showing connectivity for folded lysine residues.

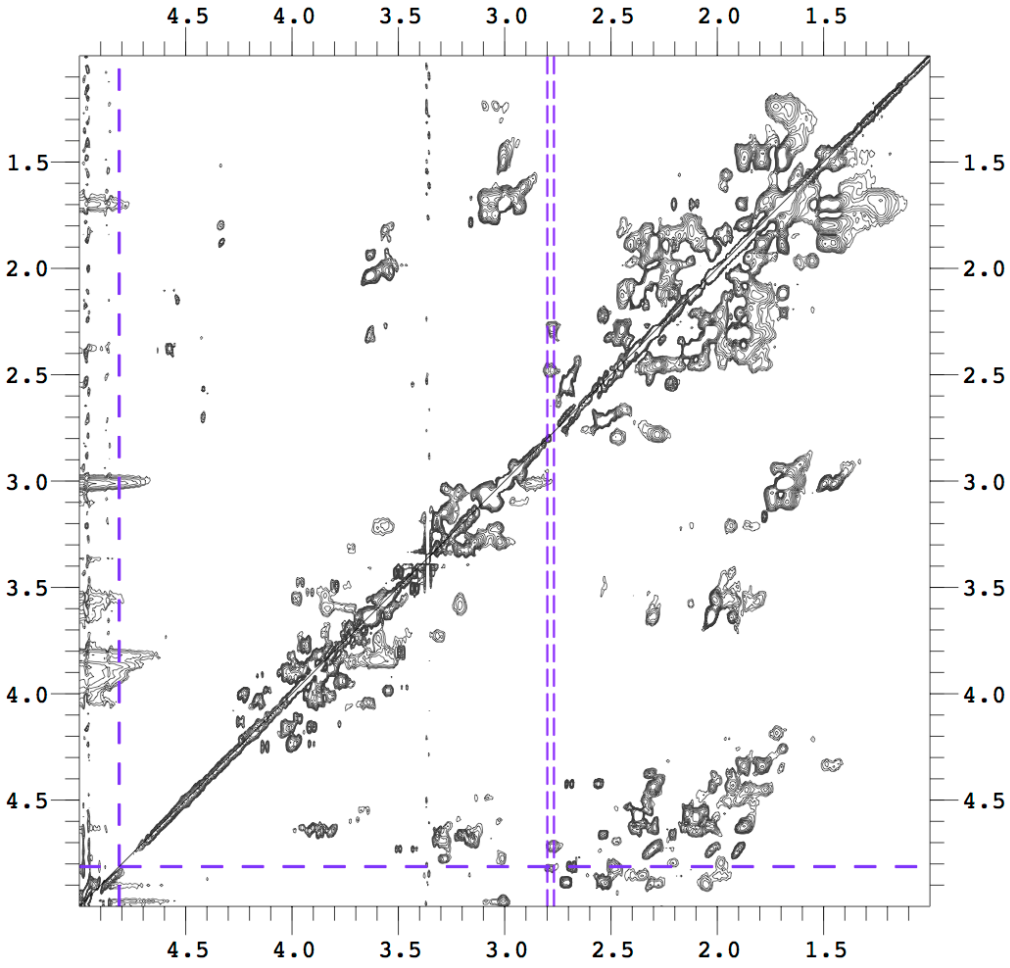


Figure A2.11. TOCSY figure showing connectivity for folded aspartate residues.

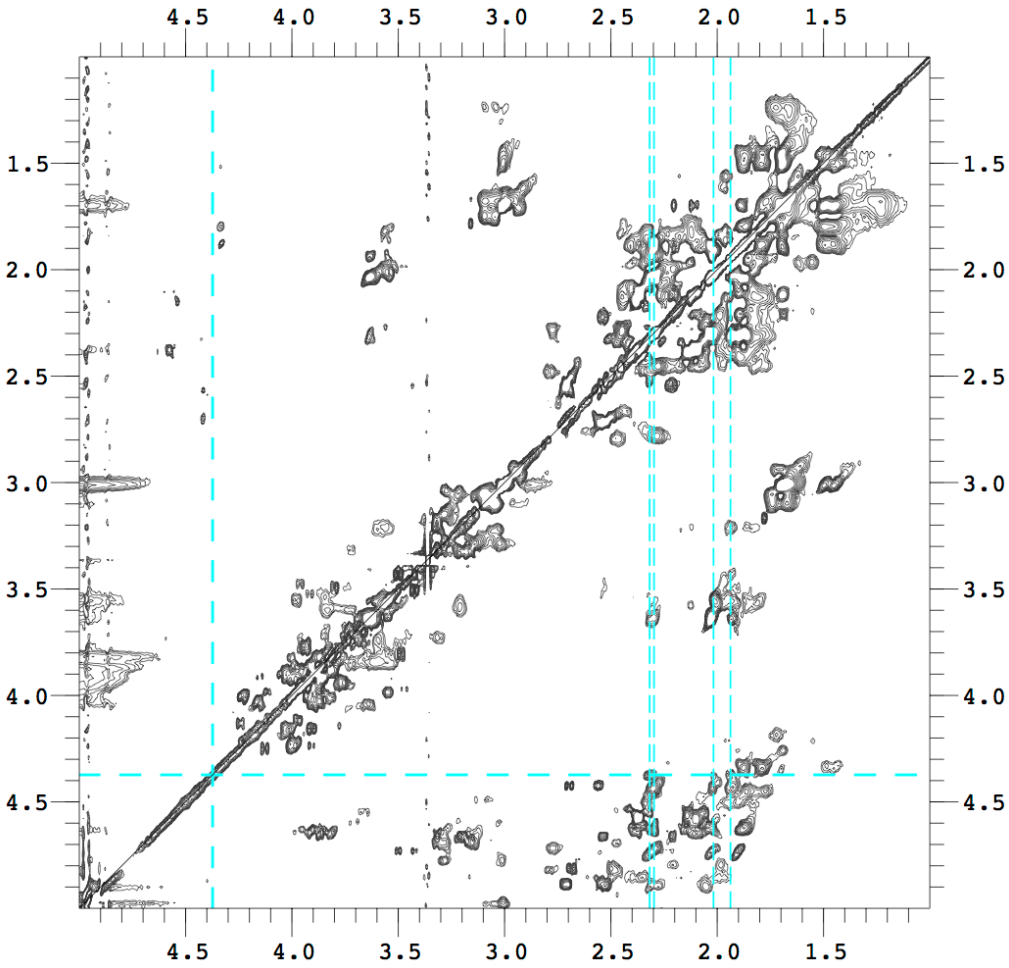


Figure A2.12. TOCSY figure showing connectivity for folded glutamate residues.

A2.3 Chapter 4

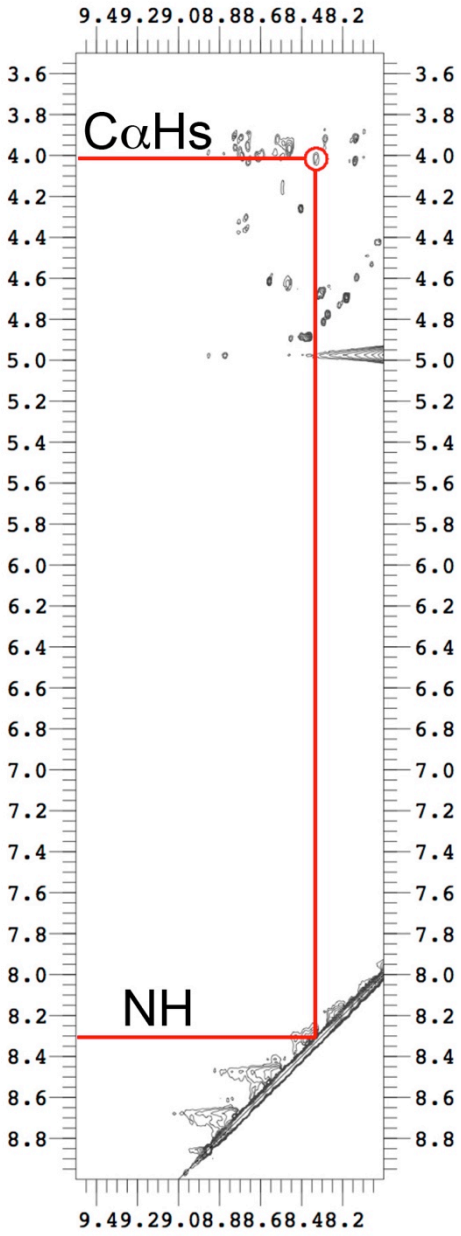


Figure A2.13. TOCSY figure of the AAB base peptide showing connectivity for Gly3 in the Base system. Chemical shifts are shown by red lines, and relevant correlations are circled.

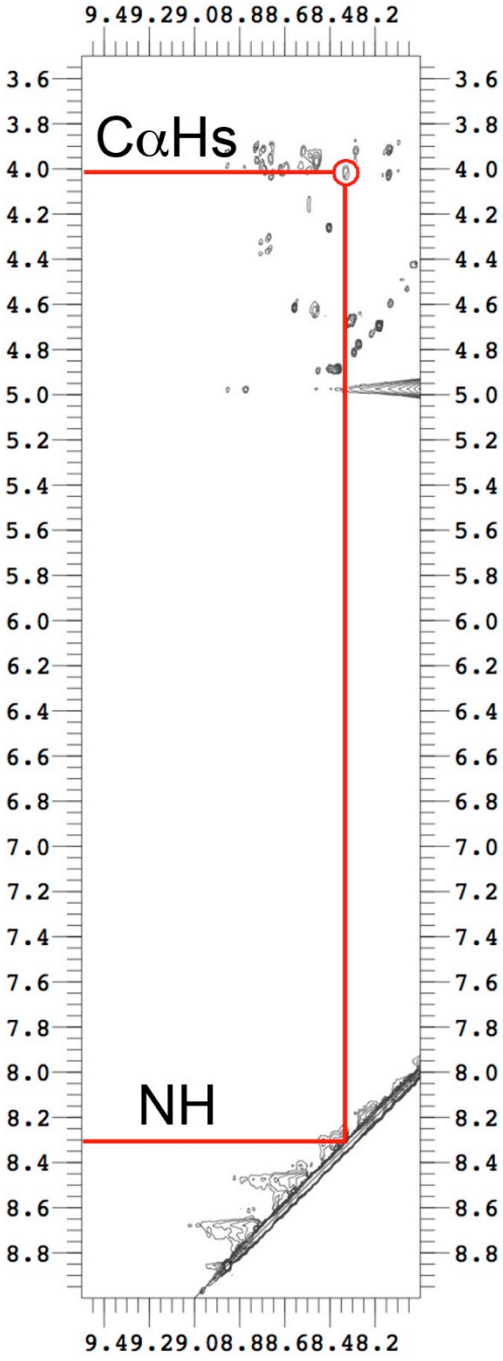


Figure A2.14. TOCSY figure of the AAB base peptide showing connectivity for Gly3 in the B-Ala system. Chemical shifts are shown by red lines, and relevant correlations are circled.

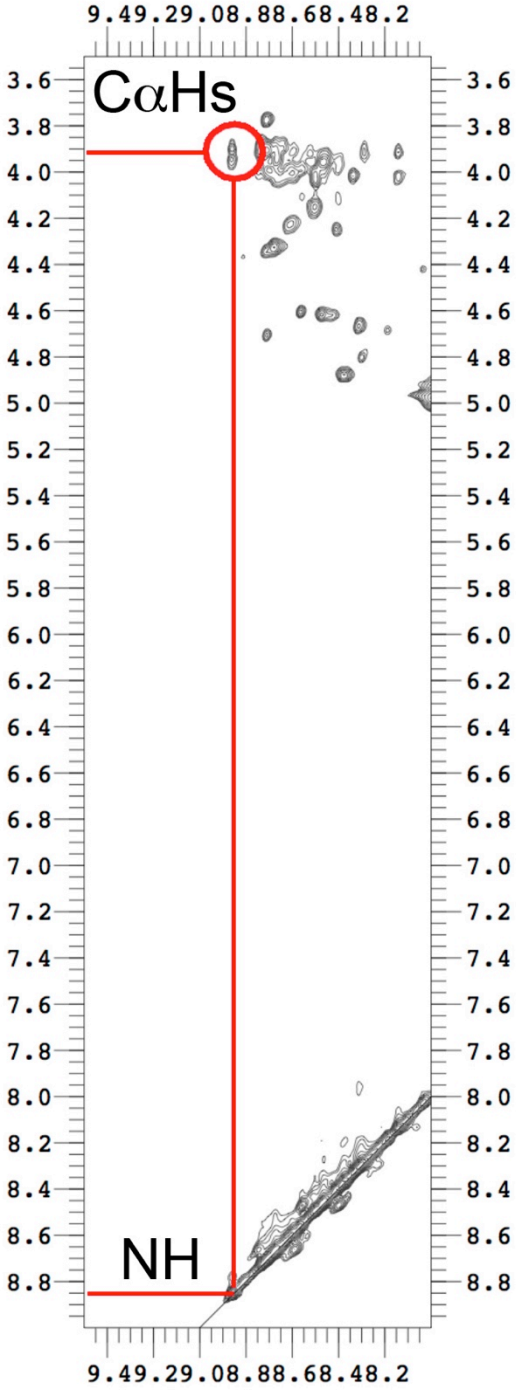


Figure A2.15. TOCSY figure of the AAB base peptide showing connectivity for Gly3 in the B-Ser system. Chemical shifts are shown by red lines, and relevant correlations are circled.

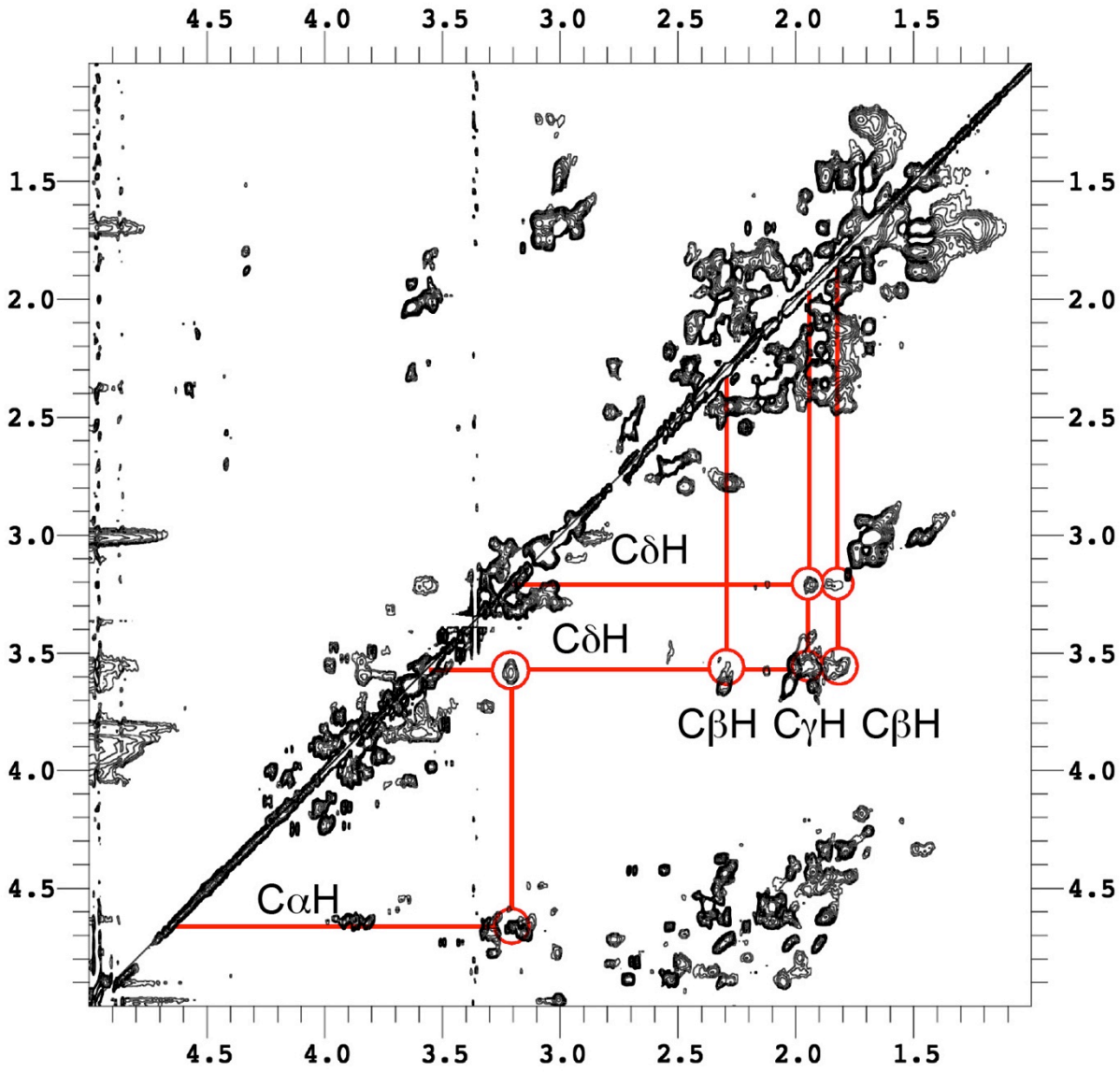


Figure A2.16. TOCSY figure of the AAB base peptide showing connectivity for folded proline residues. Chemical shifts are shown by red lines, and relevant correlations are circled. With the mixing time used during acquisition (85 ms) it was not possible to see direct CβH to CαH correlations- for shorter mixing times (25 ms) this cross peak was observed though overlapped with other peaks in the alpha proton region. Though not directly coupled, with the longer 85 ms mixing time it was possible to observe a CδH to CαH cross peak.

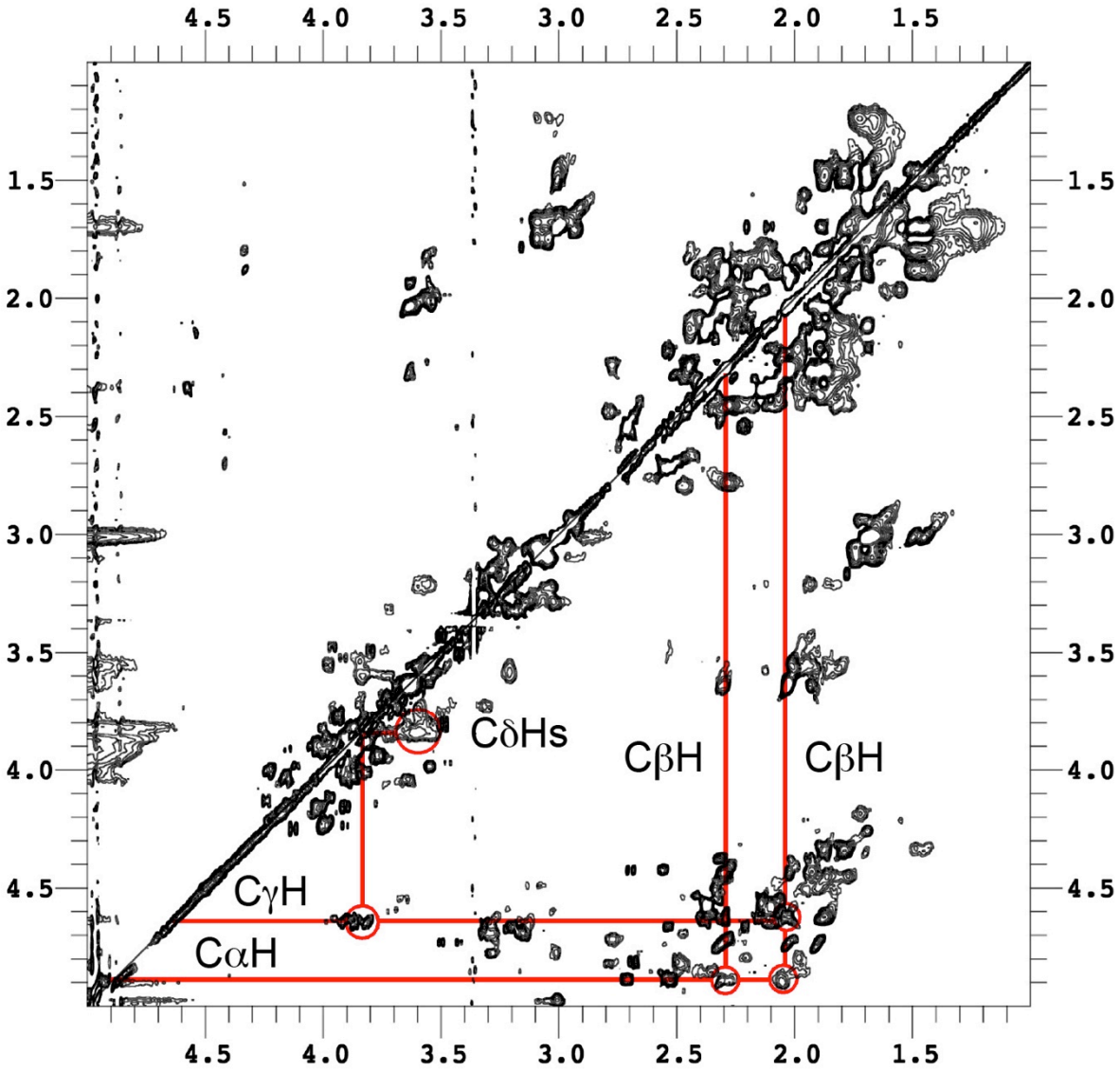


Figure A2.17. TOCSY figure of the AAB base peptide showing connectivity for folded hydroxyproline residues. Chemical shifts are shown by red lines, and relevant correlations are circled.

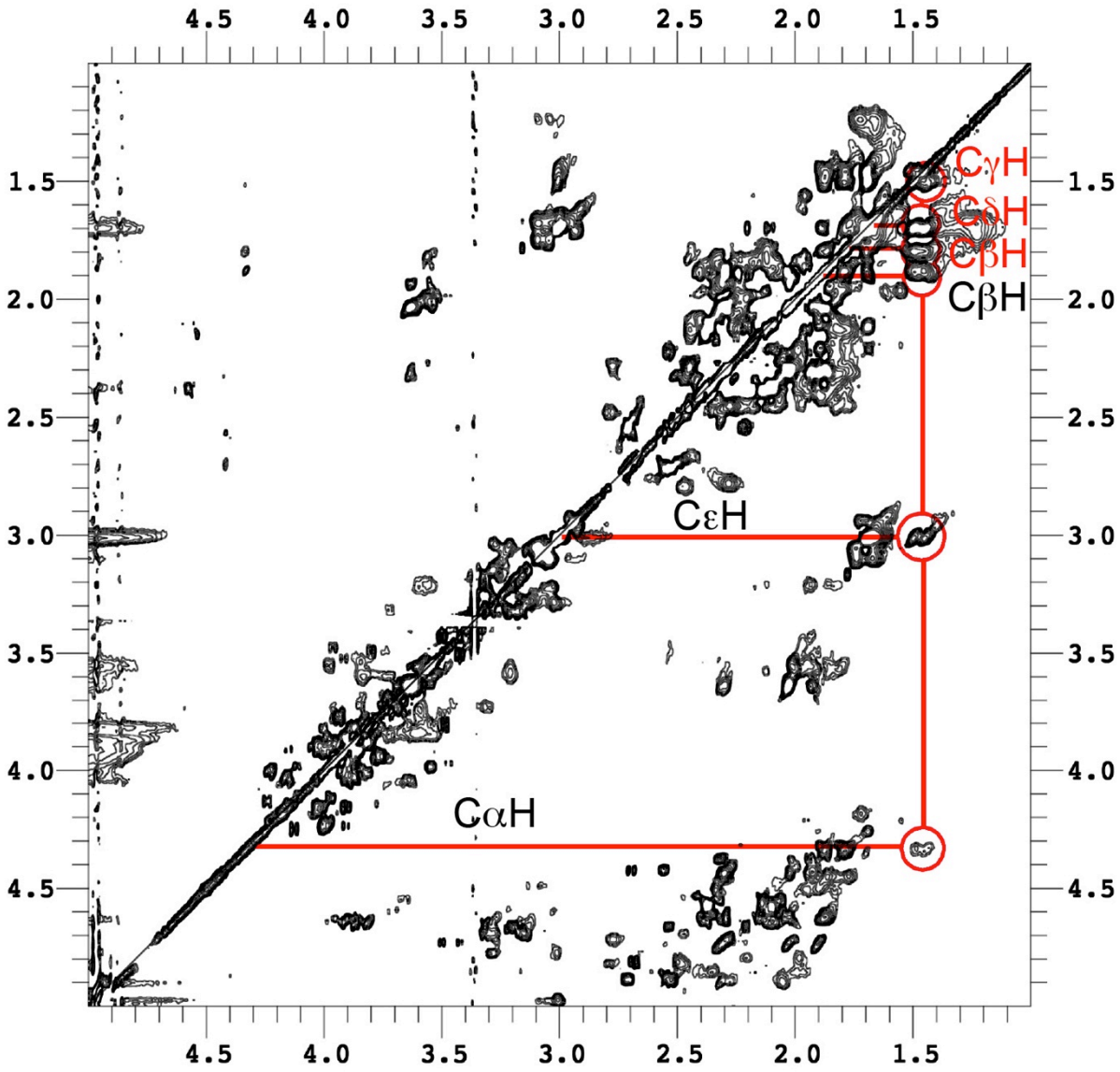


Figure A2.18. TOCSY figure of the AAB base peptide showing connectivity for folded lysine residues. Chemical shifts are shown by red lines, and relevant correlations are circled.

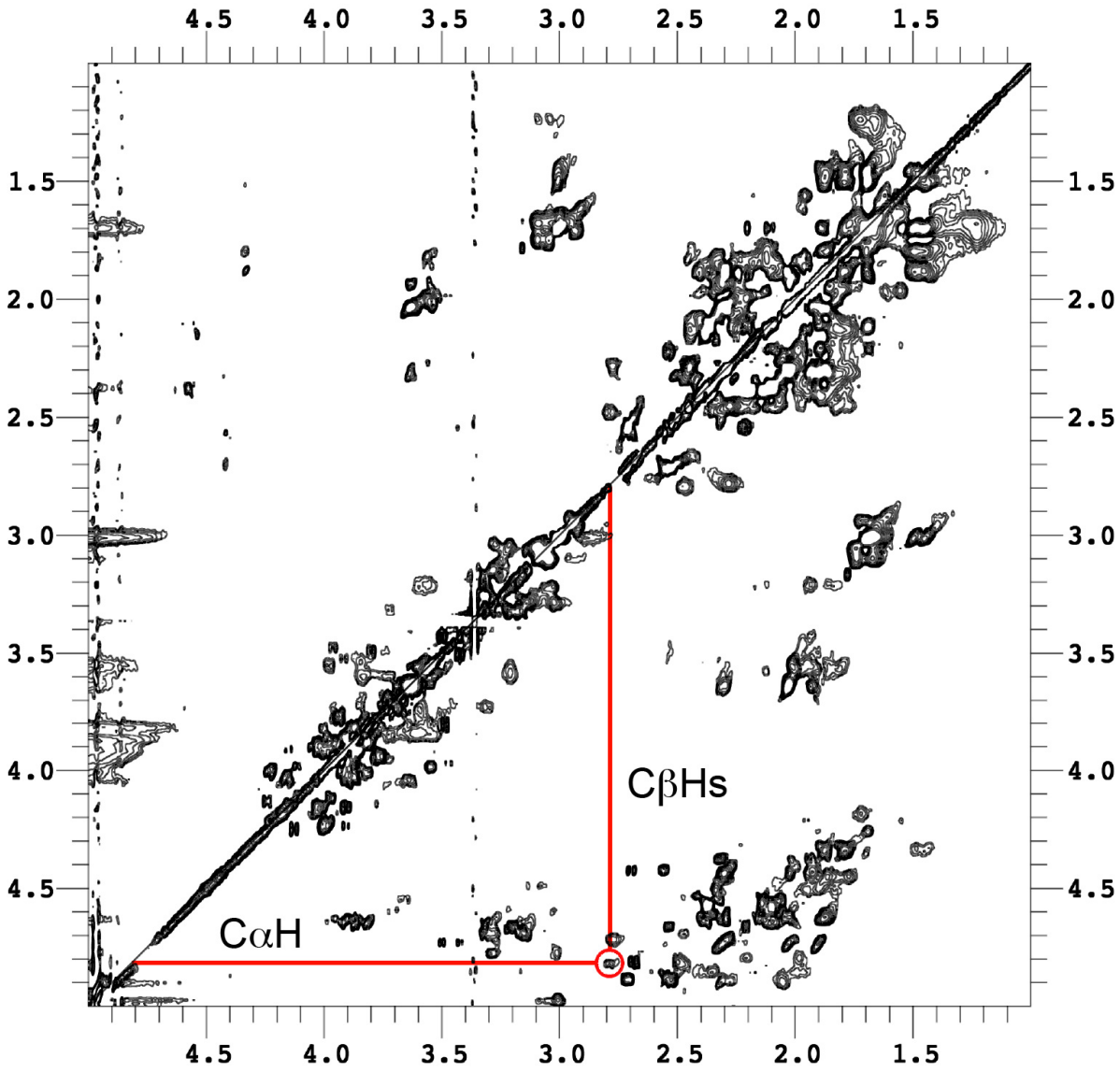


Figure A2.19. TOCSY figure of the AAB base peptide showing connectivity for folded aspartate residues. Chemical shifts are shown by red lines, and relevant correlations are circled.

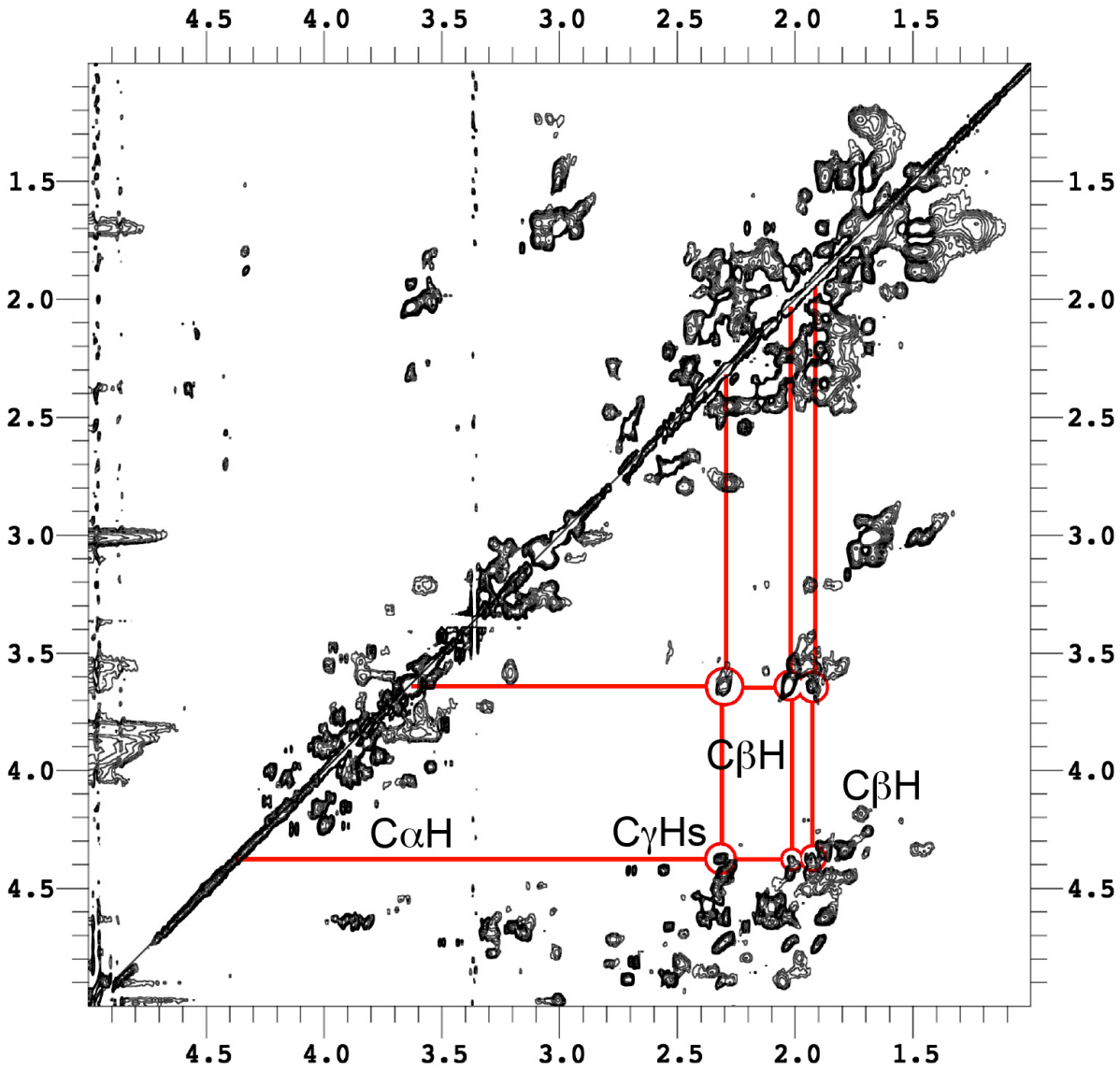


Figure A2.20. TOCSY figure of the AAB base peptide showing connectivity for folded glutamate residues. Chemical shifts are shown by red lines, and relevant correlations are circled.

Appendix 3: List of Publications

(1) Jalan, A.A.; Jochim, K.A.; Hartgerink, J.D. *J. Am. Chem. Soc*. **2014,** *136*, 7535−7538

(2) Acevedo-Jake, A.M.; Clements, K.A.; Hartgerink, J.D. *Biomacromolecules* **2016**, *17*, 914−921

(3) Clements, K.A.; Acevedo-Jake, A.M.; Walker, D.R.; Hartgerink, J.D. *Biomacromolecules* **2017**, *18*, 617-624.